AI LAW REVIEW

AI Law

International Review of Artificial Intelligence Law

1/2026

Giappichelli

AI LAW REVIEW

AI Law

International Review of Artificial Intelligence Law

1/2026

AI Law - International Review of Artificial Intelligence Law
Rivista fondata dallo Studio legale SLGLAW e dall'Avv. Marco Perilli
Rivista semestrale - Iscrizione al R.O.C. n. 25223

© Copyright - G. GIAPPICHELLI EDITORE - TORINO
VIA PO, 21
www.reviewofailaw.com

Board of Directors:

Fabio Addis, Sapienza Università di Roma (chair); **Cristina Alessi**, Università di Brescia; **Francesco Astone**, Università di Foggia; **Maria Luisa Atienza Navarro**, Universitat de València; **Sergio Avila Negri**, Universidade Federal de Juiz de Fora; **Francesca Bartolini**, Università Link Campus; **Maria Esmeralda Bucalo**, Università di Palermo; **Giuseppe Corasaniti**, Università di Brescia; **Lucia Corso**, Università di Enna "Kore"; **Floriana Cursi**, Università di Roma Tor Vergata; **Virgilio D'Antonio**, Università di Salerno; **Jose Luiz de Moura Faleiros Junior**, Pontificia Universidade Catolica do Parana; **David Fernandez De Retana Gorostiza**, Universidad de Deusto – Bilbao; **Raffaele Di Raimo**, Università Roma Tre; **Ernesto Fabiani**, Università del Sannio; **Pasquale Femia**, Università di Salerno; **Camilla Ferrari**, Università di Milano; **Giuseppe Ferri**, Sapienza Università di Roma; **Roberto Fiori**, Università di Roma Tor Vergata; **Andrea Genovese**, Università della Tuscia; **Gregorio Gitti**, Università di Milano; **Carlos Agurto Gonzales**, Unversidad Nacional Mayor de San Marcos; **Mauro Grondona**, Università di Genova; **Antoine Harfouche**, Università della Tuscia; **Inmaculada Herbosa Martínez**, Universidad de Deusto – Bilbao; **Leopoldo López Máñez**, Cuerpo Consular de València; **Nadia Maccabiani**, Università di Brescia; **Anselmo María Martínez Cañellas**, Universitat de les Illes Balears; **Michel Martone**, Sapienza Università di Roma; **Valeria Mastroiacovo**, Università di Foggia; **Enrico Minervini**, Università di Napoli Federico II; **Gaetano Natullo**, Università del Sannio; **Mauro Pennasilico**, Università di Bari Aldo Moro; **Valerio Pescatore**, Sapienza Università di Roma; **Giovanni Pesce**, Università Telematica Pegaso; **Angela Procaccino**, Università di Foggia; **Massimo Proto**, Università Link Campus; **Francesco Ricci**, Università LUM Giuseppe Degennaro; **Thomas Rüfner**, Universität Trier; **Gianpaolo Maria Ruotolo**, Università di Foggia; **Antonio Saccoccio**, Sapienza Università di Roma; **Maddalena Semeraro**, Università "Magna Graecia" di Catanzaro; **Alessandro Sterpa**, Università della Tuscia; **Jean-Marc Van Gyseghem**, Université de Namur; **Daniele Vattermoli**, Sapienza Università di Roma; **Alberto Venturelli**, Università di Brescia; **Vincenzo Verdicchio**, Università del Sannio; **Massimo Zaccheo**, Sapienza Università di Roma.

Editorial Committee:

Flavia Accoto, Università LUM Giuseppe Degennaro; **Anna Bettoni**, Università della Tuscia; **Simona Cacace**, Università di Brescia;

Claudia Capasso, Università della Tuscia; **Vincenza Conte**, Università di Salerno; **Luca Esposito**, Università di Salerno; **Luigi Follieri**, Università di Foggia; **Violetta Gargani**, Università della Tuscia; **Vera Iuzzolino**, Università di Salerno; **Giorgio Lanzetta**, Sapienza Università di Roma; **Donato Maria Matera**, Università di Brescia; **Serena Mirabello**, Sapienza Università di Roma; **Nicola Ruccia**, Università del Sannio; **Ciro Maria Ruocco**, Università di Salerno; **Marco Scaldaferri**, Sapienza Università di Roma; **Andrea Sardini**, Università di Milano; **Anna Maria Siniscalchi**, Università di Cagliari; **Valentina Viti**, Università di Brescia.

AI Law Managing Director:

Marco Perilli

AI LAW
INTERNATIONAL REVIEW
OF ARTIFICIAL INTELLIGENCE LAW

Indice

Saggi

M. FLORIANA CURSI, IA: modelli tradizionali di responsabilità e nuove prospettive 1

VINCENZO VERDICCHIO, Law, Legal Science and New Technologies 13

FABIO ADDIS, El impacto de la inteligencia artificial en la producción científica 35

ANTONELLA D'ANDREA, Judiciary, Artificial Intelligence and Human Rights 53

NICOLA SCERBO, L'*A.I.* e il suo possibile supporto nel processo di valutazione del nesso di causalità nel giudizio di colpevolezza 69

MICHAEL LOGNOUL, EU law seeks to tackle AI opacity, but can it? 101

NADIA MACCABIANI, Streamlining EU Digital Regulation While Enforcing Human-Centric Constitutional Safeguards 115

ANTONIO FELICE URICCHIO-CLAUDIO CALDAROLA, Verso un ecosistema digitale europeo integrato. Digital Omnibus, Data Act, AI Act e tutela della persona umana nella governance dei dati e dell'Intelligenza Artificiale 146

RACHELE RANIERI, La nuova realtà della giustizia predittiva 170

MIRKO FACCIOLI, La responsabilità civile del medico utilizzatore di sistemi di intelligenza artificiale nell'ordinamento giuridico italiano 211

VALERIO PESCATORE, Ethical Codes in the European Artificial Intelligence Regulation 239

AI LAW

INTERNATIONAL REVIEW
OF ARTIFICIAL INTELLIGENCE LAW

IA: modelli tradizionali di responsabilità e nuove prospettive *

M. Floriana Cursi

Professoressa ordinaria di Diritto romano e diritti dell'antichità nell'Università degli Studi di Roma Tor Vergata

Abstract

(EN): The essay explores the issue of civil liability for damages caused by artificial intelligence in light of Roman legal tradition and its subsequent doctrinal developments. Starting from the frequent comparison between the "intelligent machine" and the *servus* of Roman law, the author highlights its inadequacy: the status of the slave, a person subject to *dominica potestas*, cannot serve as a meaningful parallel for attributing liability to artificial entities. Instead, the analysis identifies in Roman law the origins of strict liability, constructed through an absolute presumption of *culpa in vigilando* or *in eligendo* imposed on the person responsible for supervising others. This model – rooted in the inseparable link between *iniuria* and *culpa* – differs fundamentally from modern vicarious liability, which, following the natural law theories of Thomasius and Pothier, introduced a relative presumption of fault and thus a subjective form of liability. The essay further examines contemporary European legislative initiatives, particularly the *AI Liability Directive*, which tends to objectify liability while maintaining the traditional fault-based structure, mainly by easing the burden of proof on the injured party. Finally, the author discusses the proposal to attribute "electronic personality" to intelligent machines, underscoring its largely symbolic rather than substantive character. The study concludes that modern trends toward the objectification of liability, though differently framed, revive the original Roman rationale of protecting the injured party by assigning responsibility to the subject best positioned to bear the cost of harm.

(ES): El ensayo examina la cuestión de la responsabilidad civil por los daños causados por la inteligencia artificial a la luz de la tradición jurídica romana y de sus desarrollos dogmáticos posteriores. Partiendo de la comparación, frecuente en la literatura, entre la "máquina inteligente" y el *servus* del derecho romano, la autora pone de relieve su inadecuación: la condición del esclavo, persona sometida a la *dominica potestas*, no permite establecer un paralelismo útil para la atribución de responsabilidad a las entidades artificiales. El análisis identifica, en cambio, en el derecho romano los orígenes de la responsabilidad objetiva, construida mediante una presunción

* Il testo riproduce la relazione tenuta nel convegno su *Intelligenza artificiale e responsabilità umana. Uno spazio non libero* presso l'Università degli Studi di Roma Tor Vergata il 7 maggio 2024.

absoluta de *culpa in vigilando* o *in eligendo* impuesta al sujeto encargado de la supervisión de otros. Este modelo – basado en el vínculo inseparable entre *iniuria* y *culpa* – se diferencia profundamente de la responsabilidad vicaria moderna que, a partir de las elaboraciones del iusnaturalismo de Thomasius y Pothier, introduce una presunción relativa de culpa y, por tanto, una forma subjetiva de responsabilidad. El estudio examina también las iniciativas legislativas europeas más recientes, en particular la *AI Liability Directive*, que tienden a objetivar la responsabilidad, aunque manteniendo la estructura tradicional basada en la culpa, principalmente mediante la atenuación de la carga de la prueba del perjudicado. Finalmente, la autora analiza la propuesta de atribuir una "personalidad electrónica" a las máquinas inteligentes, subrayando su carácter más simbólico que real. El trabajo concluye que las tendencias contemporáneas hacia la objetivación de la responsabilidad, aunque adoptan nuevas formas, reavivan la lógica romana originaria de tutela del perjudicado mediante la identificación de un sujeto responsable capaz de asumir el coste del daño.

(FR): L'étude examine la question de la responsabilité civile pour les dommages causés par l'intelligence artificielle à la lumière de la tradition juridique romaine et de ses développements doctrinaux ultérieurs. Partant de la comparaison, fréquente dans la littérature, entre la «machine intelligente» et le *servus* du droit romain, l'auteure en souligne l'inadéquation : la condition de l'esclave, personne soumise à la *dominica potestas*, ne permet pas d'établir un parallèle pertinent pour l'attribution de la responsabilité aux entités artificielles. L'analyse identifie plutôt dans le droit romain les origines de la responsabilité objective, élaborée à travers une présomption absolue de *culpa in vigilando* ou *in eligendo* imposée au sujet chargé de surveiller autrui. Ce modèle – fondé sur le lien indissoluble entre *iniuria* et *culpa* – se distingue profondément de la responsabilité vicariale moderne qui, à partir des théories jusnaturalistes de Thomasius et de Pothier, introduit une présomption relative de faute et, par conséquent, une forme subjective de responsabilité. L'étude aborde également les initiatives législatives européennes récentes, en particulier la *AI Liability Directive*, qui tendent à objectiver la responsabilité tout en maintenant la structure traditionnelle fondée sur la faute, notamment par l'allègement de la charge de la preuve du lésé. Enfin, l'auteure discute la proposition d'attribuer une «personnalité électronique» aux machines intelligentes, en soulignant son caractère davantage symbolique que réel. L'étude conclut que les tendances contemporaines vers l'objectivation de la responsabilité, bien qu'elles se présentent sous des formes nouvelles, réactivent la logique romaine originelle de protection du lésé par l'identification d'un sujet responsable apte à supporter le coût du dommage.

(DE): Der Aufsatz untersucht die Frage der zivilrechtlichen Haftung für Schäden, die durch künstliche Intelligenz verursacht werden, im Lichte der römischen Rechtstradition und ihrer späteren dogmatischen Entwicklungen. Ausgehend vom häufig gezogenen Vergleich zwischen der „intelligenten Maschine" und dem *servus* des römischen Rechts zeigt die Autorin dessen Unangemessenheit auf: Der Status des Sklaven, einer Person, die der *dominica potestas* untersteht, erlaubt es nicht, eine sinnvolle Parallele zur Zurechnung der Haftung künstlichen Entitäten gegenüber herzustellen. Stattdessen identifiziert die Analyse im römischen Recht die Ursprünge der Gefährdungshaftung, die auf einer absoluten Vermutung der *culpa in vigilando* oder *in eligendo* des Aufsichtsverpflichteten beruht. Dieses Modell – gegründet auf dem unauflöslichen Zusammenhang zwischen iniuria und culpa – unterscheidet sich grundlegend von der modernen Haftung für Dritte, die, ausgehend von den naturrechtlichen Theorien von Thomasius und Pothier, eine relative Verschuldensvermutung und damit eine subjektive Form der Haftung eingeführt hat. Die Studie befasst sich ferner mit den jüngsten europäischen Gesetzesinitiativen, insbesondere mit der *AI Liability Directive*, die dazu neigen, die Haftung zu objektivieren, dabei jedoch die traditionelle, auf Verschulden beruhende Struktur beibehalten – vor allem durch die Erleichterung der Beweislast für die geschädigte Partei. Schließlich erörtert die Autorin den Vorschlag, intelligenten Maschinen

eine „elektronische Persönlichkeit“ zuzuschreiben, und hebt deren eher symbolischen als tatsächlichen Charakter hervor. Die Arbeit schließt mit der Feststellung, dass die modernen Tendenzen zur Objektivierung der Haftung – wenn auch in neuer Form – die ursprüngliche römische Logik des Opferschutzes wiederbeleben, indem sie die Verantwortung demjenigen Subjekt zuweisen, das am besten in der Lage ist, die Kosten des Schadens zu tragen.

(PT): O ensaio analisa a questão da responsabilidade civil pelos danos causados pela inteligência artificial à luz da tradição jurídica romana e dos seus desenvolvimentos dogmáticos posteriores. Partindo da comparação, frequentemente presente na literatura, entre a “máquina inteligente” e o *servus* do direito romano, a autora evidencia a sua inadequação: a condição do escravo, pessoa sujeita à dominica potestas, não permite estabelecer um paralelo útil para a atribuição de responsabilidade às entidades artificiais. A análise identifica, pelo contrário, no direito romano as origens da responsabilidade objetiva, construída mediante uma presunção absoluta *de culpa in vigilando* ou *in eligendo* imposta ao sujeito encarregado da supervisão de outros. Esse modelo – fundado no vínculo indissociável entre *iniuria* e *culpa* – distingue-se profundamente da responsabilidade vicária moderna, que, a partir das teorias jusnaturalistas de Thomasius e Pothier, introduz uma presunção relativa de culpa e, portanto, uma forma subjetiva de responsabilidade. O estudo examina ainda as iniciativas legislativas europeias mais recentes, em particular a *AI Liability Directive*, que tendem a objetivar a responsabilidade, mantendo contudo a estrutura tradicional baseada na culpa, sobretudo mediante o alívio do ónus da prova do lesado. Por fim, a autora discute a proposta de atribuir uma “personalidade eletrónica” às máquinas inteligentes, sublinhando o seu caráter mais simbólico do que real. O trabalho conclui que as tendências contemporâneas de objetivação da responsabilidade, ainda que em novas formas, reavivam a lógica romana originária de proteção do lesado através da identificação de um sujeito responsável capaz de suportar o custo do dano.

(IT): Il saggio indaga il problema della responsabilità civile per i danni causati dall’intelligenza artificiale alla luce della tradizione giuridica romana e delle successive evoluzioni dogmatiche. Muovendo dal frequente confronto tra la “macchina intelligente” e il *servus* romano, l’autrice ne evidenzia l’inadeguatezza: la condizione dello schiavo, persona sottoposta alla *dominica potestas*, non consente di fondare un parallelo utile per la responsabilità delle entità artificiali. Il contributo individua invece nel diritto romano le origini della responsabilità oggettiva, costruita mediante la presunzione assoluta di *culpa in vigilando* o *in eligendo* a carico del soggetto tenuto al controllo di altri. Tale modello, fondato sull’inscindibile nesso tra *iniuria* e *culpa*, si differenzia radicalmente dalla responsabilità vicaria moderna, che, attraverso l’elaborazione giusnaturalista di Thomasius e Pothier, introduce una presunzione relativa e quindi una forma soggettiva di colpa. La riflessione si estende alle più recenti proposte normative europee, in particolare alla *AI Liability Directive*, che tende a oggettivizzare la responsabilità pur mantenendo la struttura soggettiva del modello colposo, attenuando l’onere probatorio del danneggiato. Infine, l’autrice discute la proposta di riconoscere una “personalità elettronica” alle macchine intelligenti, rilevandone il carattere più simbolico che reale. Il saggio conclude che le moderne tendenze alla oggettivizzazione della responsabilità ripropongono, seppure in forme nuove, la logica romana di tutela del danneggiato mediante l’individuazione di un soggetto responsabile in grado di sostenere il costo del danno.

Sommario: 1. Il superamento del confronto tra macchina intelligente e *servus* romano. – 2. La costruzione della responsabilità oggettiva nel diritto romano. – 3. Il paradigma della moderna responsabilità oggettiva e gli strumenti per alleggerire l'onere della prova del danneggiato. – 4. La responsabilità oggettiva della "persona elettronica". – 5. Corsi e ricorsi storici in tema di responsabilità oggettiva.

1. Il superamento del confronto tra macchina intelligente e *servus* romano

Negli studi relativi ai sistemi di IA, come ad esempio i sistemi robotici evoluti capaci di emulare autonomamente le facoltà proprie della persona (ragionamento, emozioni, comportamenti, decisioni), spesso si trova richiamata la figura del *servus*, considerato dai Romani contemporaneamente *persona* e *res*. In considerazione di questo peculiare statuto ontologico, la figura dello schiavo è usata quale termine di confronto per la costruzione dello *status* e del relativo regime della responsabilità da algoritmo di nuove entità intelligenti [1].

Da studiosa di diritto romano, il confronto mi ha sorpreso non poco: se ci si ferma all'apparenza, sicuramente lo schiavo compendia i due *status* di *persona* e *res*, come potrebbe dirsi della macchina intelligente, sicuramente cosa, ma capace di interagire nel contesto umano con comportamenti che la avvicinano a un individuo. Se scendiamo oltre la superficie del fenomeno, però, il confronto rileva la sua fragilità: lo schiavo romano è ontologicamente una persona che i giuristi romani riconducono alla disciplina giuridica delle *res*.

Non è un caso che, nel momento in cui lo schiavo causa un illecito, il *domi-*

[1] Uno dei primi interventi in questo senso è di Luciano Floridi, professore di filosofia ed etica dell'informazione all'Università di Oxford, in un articolo pubblicato nel 2017 sul *Financial Times*, intitolato *Roman law offers a better guide to robot rights than sci-fi*, ove instaura la connessione tra *servus*, dipendente dal *dominus*, e i *robots*, criticando la scelta del Parlamento europeo che in una Risoluzione del 2017, invitava la Commissione europea ad attribuire ad alcuni *robots* «a 'specific legal status' as 'electronic persons responsable'». Cfr. anche U. RUFFOLO, *La 'personalità elettronica'*, in ID. (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, Milano, 2020, p. 213 ss.; e ID., *VI lezione: la personalità elettronica. La personalità elettronica tra 'doveri' e 'diritti' della machina*, in ID. (a cura di), *XXVI lezioni di Diritto dell'Intelligenza Artificiale*, Torino, 2021, p. 115 s., rispetto alla personalizzazione dello schiavo impiegato nella gestione di attività economiche per conto del proprio *dominus*. Anche tra gli studiosi del diritto romano viene proposta la comparazione tra macchina intelligente e *servus* basata sul confronto tipologico delle due entità: cfr. L. FRANCHINI, *Disciplina romana della schiavitù ed intelligenza artificiale odierna. Spunti di comparazione*, in *Dir. merc. tecn*., 2020, *on line*; M. GEMELLI, *Il robot come* servus novus*: spunti di comparazione e tentativi storici di approccio alla disciplina della odierna intelligenza artificiale*, in *Camm. dir.*, 6, 2023, *on line*.

nus ne risponde solo se il *servus* abbia agito con *dolus* o *culpa*, criteri soggettivi di responsabilità della persona, scegliendo se pagare la pena o consegnare a nossa lo schiavo colpevole all'offeso. Analogamente, rispetto ai negozi compiuti dallo schiavo, per verificare se il *dominus* debba rispondere (illimitatamente o limitatamente, a seconda del contesto e dell'incarico o meno alla negoziazione) si prendono in considerazione i criteri psicologici del *servus*.

Pur essendo riconosciuto come una *persona*, lo schiavo non è però direttamente responsabile delle proprie azioni. Questo accade non tanto perché è una *persona* reificata, quanto per il fatto di essere sottoposto alla *dominica potestas* del proprio *dominus*, l'unico che può rispondere delle condotte delle persone e cose sottoposte al suo potere. È analoga, infatti, la condizione del *filius familias* che, pur essendo una persona libera e non una cosa, è comunque sottoposto alla *patria potestas* del proprio *pater*, condividendo con lo schiavo lo stesso regime giuridico: si pensi alla consegna nossale del *filius familias* in caso di illecito.

Come appare evidente, è il carattere potestativo del rapporto tra *dominus* e *servus* a qualificare la condizione dello schiavo romano. Il che obbliga, per tornare al confronto con le nuove entità intelligenti, a spostare i termini della discussione dal *servus* alla responsabilità del *dominus* per il fatto del *servus*.

2. La costruzione della responsabilità oggettiva nel diritto romano

Il modello potestativo romano, fondato sulla peculiare logica della famiglia patriarcale, non può tuttavia essere d'aiuto nel tracciare continuità con le moderne forme di responsabilità. Esso non solo è insufficiente, ma si rivela addirittura fuorviante nella ricerca del fondamento dell'attuale modello della responsabilità vicaria che, con le sue irrisolte ambiguità [2], costituisce oggi la base dogmatica delle soluzioni che gli ordinamenti europei propongono anche per i danni causati dall'intelligenza artificiale: dalla responsabilità del produttore [3], a quella della circolazione di autoveicoli, o per esercizio di attività pericolose, oppure da cose in custodia o animali, o infine del padrone e committente [4].

[2] Sul punto mi permetto di rinviare a M.F. CURSI, *Danno e responsabilità extracontrattuale nella storia del diritto privato*, II ed., Napoli, 2021, p. 359 ss.

[3] Cfr. A. AMIDEI, *Intelligenza Artificiale e responsabilità da prodotto*, in RUFFOLO (a cura di), *Intelligenza artificiale*, cit., p. 125 ss., il quale ritiene che la normativa europea in materia di *product liability*, evolutivamente interpretata o emendata legislativamente «potrebbe rivelarsi lo strumento più idoneo a dare (almeno parziale) mediazione giuridica alle istanze tecnoetiche di settore giudicate meritevoli ...».

[4] Cfr. U. RUFFOLO, *Intelligenza Artificiale ed* automotive*: responsabilità da veicoli* self-driving

Se dunque non è il modello potestativo a fondare la moderna responsabilità vicaria, dove vanno ricercate le sue radici?

Il fondamento della responsabilità per fatto altrui può rintracciarsi nella responsabilità per *culpa in vigilando* o *in eligendo* che i giuristi romani presumono in maniera assoluta, senza cioè prova liberatoria, in capo a un soggetto per i danni arrecati da terzi.

Le ipotesi sono tipizzate e si riferiscono alla responsabilità dell'*habitator* nell'*actio de effusis vel deiectis*, del *dominus*, e alternativamente dell'*inquilinus*, nell'*actio de positis vel suspensis* e, infine, dell'*exercitor* nell'*actio damni* e *furti adversus nautas, caupones, stabularios* [5].

Con le prime due azioni si tutela in un caso il danno prodotto dal contatto con qualche cosa gettata o sparsa da un'abitazione che affacci sulla strada (*actio de effusis vel deiectis*), nell'altro il mero pericolo di danno derivante dall'aver posto su una tettoia qualcosa in equilibrio precario (*actio de positis vel suspensis*), nei confronti di chi si trovi a passare per una strada di normale transito. La *ratio* delle azioni è da ravvisare nell'esigenza – resa più pressante dalla costruzione di abitazioni a più piani (*insulae*) – di rendere sicura la circolazione lungo le vie per le quali si è soliti passare. Proprio a questo scopo, la tutela non è data necessariamente contro l'autore materiale del danno, ma, nel primo caso, nei confronti dell'*habitator* della costruzione dalla quale è stato gettato o sparso qualcosa; nel secondo, nei confronti del *dominus* o dell'*inquilinus* della casa dalla quale pende qualcosa. L'eccezione alla regola della responsabilità personale si spiega, infatti, soltanto se si ritiene primaria l'esigenza di garantire la sicurezza delle strade. Nel bilanciamento di interessi tra esigenza risarcitoria e responsabilità soggettiva, la scienza giuridica romana non ha dubbi, introducendo una tutela assoluta, ispirata alla responsabilità oggettiva. Una logica analoga guida la scelta del pretore rispetto all'esercizio dell'impresa, marittima o terrestre. Qualora il danno o il furto sia realizzato dai propri dipendenti – e solo

e driverless, in ID. (a cura di), *Intelligenza Artificiale*, cit., p. 153 ss.; e ID., *La 'personalità elettronica'*, *ivi*, p. 216 ss.; ID., *VII lezione: la responsabilità da algoritmo. Le responsabilità da produzione e gestione di intelligenza artificiale* self-learning, in ID. (a cura di), *XXVI lezioni*, cit., p. 134 ss.

[5] Sul collegamento tra l'emergere della nozione di *culpa*, anche nella sua configurazione di *culpa in vigilando* e *in eligendo*, e la patrimonializzazione del danno, rinvio a M.F. CURSI, *Roman delicts and the construction of fault*, in T. MCGINN (ed.), *Obligations in Roman Law. Past, present, and future*, Ann Arbor, 2012, p. 296 ss.; EAD., *La colpa e i profili risarcitori del danno*, in Carmina iuris. *Mélanges en l'honneur de Michel Humbert*, Paris, 2012, p. 227 ss.; EAD., *Modelle objektiver Haftung im Deliktsrecht: Das schwerwiegende Erbe des römischen Rechts*, in *ZRG RA*, (132) 2015, p. 362 ss.; EAD., *The Scope and Function of Civil Wrongs in Roman Society*, in P.J. DU PLESSIS-C. ANDO-K. TUORI (ed.), *The Oxford Handbook of Roman Law and Society*, Oxford, 2016, p. 596 ss.

da questi, non dai terzi estranei all'impresa –, l'imprenditore risponde direttamente, proprio al fine di garantire integralmente gli avventori che vengano a contatto con l'attività dell'impresa [6].

Il meccanismo inventato dai giuristi romani per giustificare questa responsabilità per fatto altrui, estranea alle logiche tradizionali dell'appartenenza, è la presunzione in capo al *pater* o *dominus*, nei suoi diversi ruoli, di un comportamento colposo consistente nel non aver adeguatamente controllato, e talora scelto, i propri schiavi, i membri della propria famiglia, gli ospiti nella propria abitazione, o i propri dipendenti nell'esercizio dell'impresa.

Il rinvio all'elemento soggettivo della *culpa* non è di per sé una novità: i giuristi, infatti, hanno adattato analogicamente a queste nuove fattispecie tutelate dal pretore modelli di responsabilità personale più risalenti (si pensi alla *culpa* aquiliana). La reale novità, che porta allo snaturamento di un simile criterio, sta nel presumere in maniera assoluta, senza cioè prova liberatoria, questa riprovevolezza in capo ad un soggetto che potrebbe essere diverso dall'autore materiale del fatto illecito, rendendolo comunque responsabile, in virtù del suo ruolo. Viene da chiedersi per quale motivo i Romani abbiano scelto di snaturare lo schema tradizionale della colpa per assicurare in termini oggettivi il risarcimento del danno.

La risposta a questa domanda risiede nell'inscindibile legame che i giuristi romani instaurano tra antigiuridicità (*iniuria*) e colpevolezza (*culpa*): la condotta illecita per essere tale deve essere anche riprovevole, e viceversa. Quella che a noi oggi sembra una costruzione stravagante – la responsabilità oggettiva elaborata attraverso la colpa – per i Romani è l'unica soluzione possibile, in virtù del nesso indissolubile tra *iniuria* e *culpa* [7].

L'interrogativo che a questo punto emerge in maniera pressante è un altro: perché ancora oggi si continua a impostare la discussione sulla responsabilità per fatto altrui in termini di *culpa* (*culpa in educando*, *in vigilando* dei genitori, *culpa in vigilando* del padrone dell'animale, etc.), quando il moderno modello della responsabilità oggettiva come responsabilità senza colpa potrebbe garantire quella stessa esigenza di tutela assoluta introdotta dai Romani?

[6] Cfr. M.F. CURSI, *Danno e responsabilità extracontrattuale nella storia del diritto privato*, cit., p. 98 ss.

[7] Mi permetto di rinviare al mio *Iniuria cum damno. Antigiuridicità e colpevolezza nella storia del danno aquiliano*, Milano, 2002, *passim*.

3. Il paradigma della moderna responsabilità oggettiva e gli strumenti per alleggerire l'onere della prova del danneggiato

La ragione va ricercata nella tradizione giuridica intermedia e in particolare in due momenti: l'elaborazione da parte del giusnaturalista Christian Thomasius della moderna responsabilità oggettiva come responsabilità senza colpa; e l'interpretazione delle fonti romane sulla responsabilità vicaria che il padre del *Code Napoléon*, il giurista Robert-Joseph Pothier, ha condotto, distinguendo tra responsabilità soggettiva e responsabilità oggettiva nella nuova prospettiva giusnaturalista.

Come è noto, la riflessione dei giusnaturalisti sul danno ruota attorno al precetto del *neminem laedere*, al cui fondamento si pone l'esigenza di custodire l'*aequalitas* tra gli uomini. È proprio il richiamo al precetto di *ius naturale* che porta Thomasius – in una delle sue più famose *dissertationes*, intitolata *Larva legis Aquiliae detracta actioni de damno dato, receptae in foris Germanorum* [8] – a rinnegare l'eredità romana dell'*actio legis Aquiliae* e dunque a rifondare su nuove basi la logica risarcitoria del danno al di fuori del contratto. Il titolo dell'opera è già di per sé piuttosto eloquente: l'intera trattazione è volta a dimostrare come l'azione di danno recepita nei tribunali tedeschi non abbia più nulla a che vedere con l'antica *actio ex lege Aquilia*.

Il diretto portato di questa novità è da ravvisare nella centralità che assume nella teoria del giurista il precetto di *ius naturae vel gentium* che fissa il dovere di risarcire il danno arrecato a un terzo – imperativo dettato sia dalla *communis tranquillitas* sia dall'*aequalitas humani generis*. Più oltre, in posizione assolutamente subordinata, troviamo il principio secondo il quale «nemo laedendus est». La gerarchia tra i due precetti è impostata in termini eminentemente funzionali: sarebbe inutile, infatti, stabilire che nessuno debba essere leso se, dopo la lesione, non ci si preoccupasse di garantire un adeguato risarcimento.

Le conseguenze di una simile costruzione sono piuttosto rilevanti: privilegiare il risarcimento del danno, rispetto alla lesione che lo ha prodotto, significa concentrare l'attenzione esclusivamente sull'evento dannoso e dunque ammettere in ogni circostanza la sua risarcibilità – sia nell'ipotesi che il soggetto abbia agito con dolo o negligenza, sia nel caso in cui «mero casu alteri damnum datum sit ...» [9].

Come è evidente, questa è la via che porta direttamente al criterio della responsabilità oggettiva. Il fondamento dogmatico di una simile responsabilità è

[8] Chr. THOMASII, *Larva legis Aquiliae detracta actioni de damno dato, receptae in foris Germanorum (LX)*, in *Dissertationes academicarum*, II, Halae Magdeburgicae, 1774, p. 849 ss.

[9] Chr. THOMASII, *op. cit*., p. 850, §§ IV-V (anche se il giurista mette in rilievo la difficoltà di risarcire il *dominus* in alcuni casi di danneggiamento *casu*).

ravvisato nella nozione obiettiva di *iniuria* che appare mutuata dallo *ius gentium* ed è sempre più lontana dalla concezione della *culpa*, ed anzi – si direbbe – segue percorsi assolutamente autonomi, in contrasto con i criteri romani della responsabilità.

Il sostrato ideologico maturato nella riflessione dei giusnaturalisti, che ha lasciato emergere momenti di continuità ma anche di decisa rottura con la tradizione, costituisce la premessa per la costruzione delle esperienze codicistiche europee.

Attraverso la mediazione di Domat [10] e Pothier [11], le norme in tema di responsabilità extracontrattuale per fatto proprio e per fatto altrui si riallacciano alla tradizione romana che, come si ricorderà, era incentrata sulla concezione romana dell'identità tra *iniuria* e *culpa*. Non è un caso, quindi, che l'unico criterio al quale è subordinato il risarcimento del danno nel *Code Napoléon* sia la *faute* di colui che arreca il danno, con la precisazione che la responsabilità nasce «non seulement par son fait, mais encore par sa négligence ou par son imprudence» [12].

Nella riflessione di Pothier [13] sulla responsabilità per fatto altrui, il modello romano fondato sulla colpevolezza dell'agente viene assolutizzato in senso soggettivo per la latente contrapposizione con il criterio della nuova responsabilità oggettiva, quale responsabilità senza colpa. L'effetto di una simile lettura è la coerente introduzione della prova liberatoria in favore del danneggiante, al fine di superare la presunzione di colpa, non più assoluta come per i Romani, ma relativa. Ulteriore conseguenza è lo snaturamento della concezione romana della responsabilità oggettiva, trasformata, sulla base delle nuove categorie dogmatiche del giusnaturalismo, in una forma di responsabilità soggettiva con tanto di prova liberatoria [14].

All'esito di questa vicenda, il giurista moderno si trova dunque a fare i conti con l'eredità romana di una responsabilità vicaria costruita sulla *culpa* (*culpa in eligendo*, *in vigilando*, *in educando*) che, distinta ormai dall'*iniuria*, non si pre-

[10] J. DOMAT, *Les loix civiles dans leur ordre naturel*, in *Œuvres complètes de Jean Domat*, I, Paris, 1835, p. 470 (lib. II, tit. VIII, *Des dommages causés par des fautes qui ne vont pas à un crime ni à un délit*).

[11] R.J. POTHIER, *Traité des obligations*, in M. BUGNET (a cura di), *Oeuvres de Pothier*, II, Paris, 1848, p. 57 ss. (part. I, cap. I, sez. II, § 2, nn. 116 ss.).

[12] Art. 1383 c.c. fr. (1804).

[13] R.J. POTHIER, *op. cit.*, pp. 58 (part. I, cap. I, sez. II, § 2, n. 121); 253 (part. II, cap. VI, sez. VIII, § 5, n. 454).

[14] Per una più ampia descrizione della vicenda rinvio a M.F. CURSI, *Danno e responsabilità extracontrattuale nella storia del diritto privato*, cit., pp. 177 ss., 227 ss.

sta a quella oggettivizzazione che aveva consentito ai Romani di garantire il risarcimento del danno in fattispecie nelle quali esso non era immediatamente riconducibile alla colpevolezza dell'agente. La contrapposizione giusnaturalista tra torto soggettivo e torto oggettivo, nel configurare la moderna responsabilità oggettiva come priva di colpa, costringe l'attuale responsabilità vicaria entro i confini della soggettività, perdendo di vista l'obiettivo risarcitorio. Nel graduale scollamento tra struttura e funzione trova dunque spazio la moderna ambiguità interpretativa della responsabilità per fatto altrui.

Questo modello, come si è detto, nelle sue diverse forme, viene esteso anche alle ipotesi di danni causati dall'intelligenza artificiale. E con esso anche le difficoltà già rilevate negli ambiti di applicazione previsti dal legislatore.

Non c'è dubbio che, da un punto di vista dogmatico, la risposta migliore all'esigenza economico-sociale, prima che giuridica, di risarcire il danno nelle ipotesi di responsabilità per fatto altrui sia fornita dal modello della responsabilità oggettiva seguendo il criterio *cuius commoda eius et incommoda*, o il rischio d'impresa, oppure l'idea che il costo del danno debba gravare sul soggetto in grado di sopportarlo nel modo più economico possibile, perché si trova nella posizione più adeguata per condurre l'analisi costi-benefici, comparando i costi necessari a evitare il danno e il costo del danno stesso [15].

La responsabilità oggettiva rappresenta il traguardo mediato della proposta di Direttiva pubblicata dalla Commissione europea nel settembre 2022 sull'adeguamento delle norme in materia di responsabilità extracontrattuale ai danni causati dall'intelligenza artificiale (*AI Liability Directive*). La possibilità del ricorso a forme di responsabilità oggettiva, unitamente alla necessità di una copertura assicurativa obbligatoria [16], viene presa in considerazione nella seconda fase dell'approccio su cui è costruita la direttiva rispetto a sistemi IA con un particolare profilo di rischio. Nella prima fase, gli obiettivi di armonizzazione sono perseguiti con un intervento meno invasivo, tenendo fermo il tradizionale modello soggettivo di responsabilità. Il danneggiante risponde infatti per colpa, anche se viene previsto un alleggerimento dell'onere probatorio del danneggiato. Per un verso, il danneggiato può accedere a informazioni provenienti direttamente dal fornitore, qualora il mezzo sia un sistema IA ad alto rischio. Per l'altro, è stata introdotta una presunzione relativa del nesso causale tra l'evento – l'azione o l'omissione derivante dall'impiego dei sistemi di IA – e la colpa del danneggiante, la cui prova è più o meno rigorosa a seconda che il sistema di

[15] Il richiamo è alla nota teoria di G. CALABRESI, *The Costs of Accidents*, Yale, 1970, con gli aggiustamenti ricordati da C. CASTRONOVO, *La nuova responsabilità civile*, III ed., Milano, 2006, in particolare p. 74 ss.

[16] Indicazioni già emerse nella Risoluzione del Parlamento europeo del 16 febbraio 2017.

IA sia o meno ad alto rischio. Così facendo, si è superata la situazione di stallo derivante dalla difficoltà per il danneggiato di individuare il nesso eziologico tra l'inosservanza di un dovere di diligenza e il risultato prodotto dall'IA da cui consegua un danno. Al danneggiante è comunque riservata la facoltà di dimostrare di aver fatto tutto il possibile al fine di evitare il danno [17].

Al di là dei problemi di allocazione della responsabilità (programmatore, produttore, operatore), questa nuova responsabilità si ispira evidentemente al modello tradizionale della responsabilità per colpa, talora presunta, in ogni caso temperata da una presunzione (superabile) del nesso causale tra colpa e danno. È l'introduzione della presunzione che, come per la responsabilità vicaria, induce a domandarsi se siamo di fronte a una tradizionale responsabilità per colpa, o se sotto questa apparenza si cela una forma oggettiva di responsabilità, lambita di fatto dal meccanismo delle presunzioni.

4. La responsabilità oggettiva della "persona elettronica"

L'alternativa alla responsabilità vicaria è riconoscere alla macchina intelligente una "personalità elettronica", dotandola di patrimonio con il quale possa rispondere direttamente dei danni arrecati a terzi. Questa è la proposta di Gunther Teubner [18], il quale affronta il tema della soggettività di agenti *software* autonomi, proponendo una capacità giuridica parziale, a metà tra il mero strumento e il soggetto giuridico di pieno diritto, che li renda responsabili oggettivamente (responsabilità da assistenza digitale) [19].

Un orientamento, quest'ultimo, fondato sulla Risoluzione del Parlamento europeo del 16 febbraio 2017 che contiene la proposta di costituire per i *robot* evoluti una innovativa finzione giuridica, attribuendo lo *status* di «persone elettroniche responsabili» a determinati sistemi di intelligenza artificiale.

L'impostazione presuppone una rivoluzione nel rapporto naturalistico per-

[17] Si legge nella Direttiva che: «la proposta non determina un'inversione dell'onere della prova per evitare di esporre i fornitori, gli operatori e gli utenti dei sistemi di IA a rischi di responsabilità più elevati, che potrebbero ostacolare l'innovazione e limitare l'adozione di prodotti e servizi basati sull'IA».

[18] G. TEUBNER, *Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Sotfwareagenten*, in *AcP*, (218) 2018, p. 155 ss.; ora in P. FEMIA (a cura di), *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, Napoli, 2019, *passim*; Cfr. anche G. ALPA, *L'intelligenza artificiale. Il contesto giuridico*, Modena, 2021, p. 139 ss.

[19] G. TEUBNER, *Soggetti giuridici digitali?*, cit., p. 92 ss. Cfr. anche P. MORO, *Macchine come noi. Natura e limiti della soggettività robotica*, in RUFFOLO (a cura di), *Intelligenza artificiale*, cit., p. 45 ss.

sona-cosa, tradotto in termini logico-giuridici di soggetto-oggetto di diritto, e implica il superamento del modello di responsabilità fondato su criteri soggettivi propri della persona umana. Ciò nonostante, l'innovativa proposta non risolve il problema di chi sia tenuto a pagare in caso di danno prodotto dalla macchina. Oltre al fatto che, attribuendo un'autonoma personalità alla macchina intelligente, si potrebbe correre il rischio di deresponsabilizzare chi poteva essere chiamato a rispondere vicariamene, con l'effetto di vedere limitato il risarcimento al fondo patrimoniale attribuito alla macchina intelligente [20].

Insomma, questa soluzione sembra essere più apparente che reale, essendo fondata su una costruzione, giudicata da taluni sostanzialmente inutile [21], atteso che la possibilità di costituire un fondo riservato al risarcimento dei danni delle macchine intelligenti può prescindere dall'attribuzione di una specifica soggettività.

5. Corsi e ricorsi storici in tema di responsabilità oggettiva

Per concludere, appare evidente come le diverse soluzioni proposte in tema di responsabilità da algoritmo manifestino la tendenza a oggettivizzare la responsabilità, superando il modello tradizionale fondato sulla colpa. Nonostante la diversa costruzione teorica della moderna responsabilità oggettiva, sta emergendo la *ratio* che aveva originariamente ispirato l'invenzione romana della responsabilità oggettiva: favorire il danneggiato individuando un soggetto responsabile sul quale far ricadere il costo del danno.

Per quanto si è detto, dunque, il confronto tra l'esperienza antica e quella moderna non può esaurirsi nella comparazione tipologica tra macchina intelligente e *servus*: esso si svolge bensì lungo la complessa vicenda che vede il modello romano della responsabilità oggettiva, costruito su una presunzione assoluta di colpa, snaturato nella moderna responsabilità vicaria, impostata sulla presunzione relativa di colpa, e solo di recente recuperato dalla tendenza a oggettivizzare la responsabilità per favorire il danneggiato, assicurandogli il risarcimento del danno.

[20] U. RUFFOLO, *La 'personalità elettronica'*, cit., p. 213 ss.; ID., *VI lezione*, cit., p. 115 ss.

[21] Cfr. G. FINOCCHIARO, *Intelligenza artificiale. Quali regole?*, Bologna, 2024, p. 66 ss.

AI LAW

INTERNATIONAL REVIEW
OF ARTIFICIAL INTELLIGENCE LAW

Law, Legal Science and New Technologies

Vincenzo Verdicchio

Professore ordinario di Diritto privato
nell'Università degli Studi del Sannio

Abstract

(EN): Given the enormous and growing proliferation of the various applications of Artificial Intelligence, and acknowledging – alongside the potential benefits – the serious risks they entail, the need for careful ethical and legal oversight is emphasized. Such oversight must be guided by the personalist principle and by full respect for human freedom and dignity. The text highlights the requirements that regulatory activity must meet in this field and the main characteristics it should possess. It also stresses the need for legal scholars to overcome the divide between scientific and humanistic cultures, in the awareness of the fundamental unity of the «two cultures».

(ES): Dada la enorme y creciente proliferación de las diversas aplicaciones de la Inteligencia Artificial, y reconociendo – junto con los posibles beneficios – los graves riesgos que conllevan, se subraya la necesidad de un control ético y jurídico riguroso. Dicho control debe inspirarse en el principio personalista y en el pleno respeto de la libertad y la dignidad humanas. El texto pone de relieve los requisitos que debe cumplir la actividad normativa en esta materia y las principales características que debe presentar. Asimismo, se destaca la necesidad de que los juristas superen la fractura entre la cultura científica y la cultura humanística, en la conciencia de la unidad fundamental de las «dos culturas».

(FR): Compte tenu de la prolifération immense et croissante des diverses applications de l'intelligence artificielle, et en reconnaissant – aux côtés des avantages potentiels – les graves risques qu'elles comportent, la nécessité d'un encadrement éthique et juridique rigoureux est soulignée. Un tel encadrement doit être guidé par le principe personnaliste et par le plein respect de la liberté et de la dignité humaines. Le texte met en lumière les exigences que l'activité normative doit satisfaire dans ce domaine, ainsi que les principales caractéristiques qu'elle doit présenter. Il souligne également la nécessité, pour les juristes, de dépasser la fracture entre culture scientifique et culture humaniste, dans la conscience de l'unité fondamentale des «deux cultures».

(DE): Angesichts der enormen und stetig wachsenden Verbreitung der verschiedenen Anwendungen der Künstlichen Intelligenz und in Anerkennung der – neben den potenziellen Vorteilen – damit verbundenen ernsthaften Risiken wird die Notwendigkeit einer sorgfältigen ethisch-rechtlichen Kontrolle betont. Eine solche Kontrolle muss vom personalistischen Prinzip und vom uneingeschränkten Respekt vor der menschlichen Freiheit und Würde geleitet sein. Der Text hebt die Anforderungen hervor, denen die gesetzgeberische Tätigkeit in diesem Bereich genügen muss, sowie die wesentlichen Merkmale, die sie aufweisen sollte. Ebenso wird die Not-

wendigkeit betont, dass Juristinnen und Juristen die Kluft zwischen naturwissenschaftlicher und geisteswissenschaftlicher Kultur überwinden – im Bewusstsein der grundlegenden Einheit der «zwei Kulturen».

(PT): Dada a enorme e crescente proliferação das diversas aplicações da Inteligência Artificial, e reconhecendo – ao lado dos potenciais benefícios – os graves riscos que elas envolvem, destaca-se a necessidade de um controle ético e jurídico rigoroso. Esse controle deve ser orientado pelo princípio personalista e pelo pleno respeito à liberdade e à dignidade humanas. O texto evidencia as exigências que a atividade normativa deve satisfazer nesse campo, bem como as principais características que deve apresentar. Ressalta-se também a necessidade de que os juristas superem a cisão entre a cultura científica e a cultura humanística, na consciência da unidade fundamental das «duas culturas».

(IT): Considerata l'enorme e crescente diffusione delle diverse applicazioni dell'Intelligenza Artificiale, e riconosciuti – accanto ai potenziali benefici – i gravi rischi che esse comportano, si sottolinea la necessità di un attento controllo etico e giuridico. Tale controllo deve essere ispirato al principio personalista e al pieno rispetto della libertà e della dignità umane. Il testo mette in evidenza le esigenze cui deve rispondere l'attività normativa in questo ambito, nonché le principali caratteristiche che essa deve presentare. Si evidenzia inoltre la necessità, per i giuristi, di superare la frattura tra cultura scientifica e cultura umanistica, nella consapevolezza dell'unità fondamentale delle «due culture».

Summary: 1. The concerns raised by the so-called Artificial Intelligence (AI). – 2. The opportunities and risks linked to AI applications (a brief inventory). – 3. The resulting need for ethical oversight of algorithms. – 4. The unavoidable subjection of digital technologies to the control of the existing legal system and the challenges of ensuring effective protection. – 5. The underlying needs of digital technology regulation and the methods and characteristics of rulemaking. – 6. Legal scholars and legal science facing the digital revolution: towards the definitive overcoming of the «two cultures».

1. The concerns raised by the so-called Artificial Intelligence (AI)

The prodigious and widespread development of new digital and information technologies and their growing impact on individuals' existential sphere as well as on the mechanisms of social interaction, has by now become evident to all, whether or not these technologies fall within the scope of the so-called Artificial Intelligence (hereinafter "AI"), depending on how one chooses to define it [1].

[1] See *infra*, note 6.

Among all the recalled technologies, AI represents, so to speak, the most advanced frontier of them, and it is from its analysis that the reflections intended here may begin.

Virtually all human activities are, in fact, susceptible to being affected by AI applications (and more broadly, by digital technologies). These technologies seem to offer unprecedented and wide-ranging opportunities –whether current, potential, or even merely imagined – for transcending (some of) the many limits that have always defined the inherent finitude of the human being. In doing so, they appear to herald the dawn of a New World [2], in which humanity redefines – on cultural, spiritual, and anthropological levels – its very representation of existential reality, its self-perception, and ultimately, its identity [3].

Ethical and philosophical reflection is thus debating whether this sweeping technological revolution is in fact overturning the classical paradigm of the human/machine relationship: that is, whether technology, having traditionally played a subordinate role as a tool – however powerful – in the service of human decision-making, is now positioning itself as a full substitute for such decisions, thereby bypassing – through its capacity for autonomous choices – its very creator.

In other words, is AI merely a new human technology, operating within a dimension that remains fundamentally anthropocentric, or are we instead facing with a technology that produces an "intelligence" destined to surpass and re-

[2] The expression «nuovo mondo» is used, referring to AI, by A. PATRONI GRIFFI, *Bioetica, diritti e intelligenza artificiale: una relazione da costruire*, in ID. (a cura di), *Bioetica, diritti e intelligenza artificiale*, Milano-Udine, 2023, p. 9, «come metafora (…) di quella scoperta [dell'America] che dal 1492 rivoluzionò quello che divenne il "vecchio mondo"»; it is indeed – the A. continues – a «straordinaria rivoluzione tecnologica capace di influire profondamente sulla vita dell'uomo e quindi sulle stesse riflessioni della bioetica, proprio nella sua accezione etimologica di etica della vita».

[3] A. PATRONI GRIFFI, *op. loc. cit.*, according to whom «Il punto ineludibile di tensione, su cui si interroga il pensiero filosofico, è se tale rivoluzione sia tanto straordinaria da mettere in discussione la stessa identità umana»; the beloved one Pope Francis also reflected on this point, in the *Address* he delivered on June 14, 2024, at the G7 session on Artificial Intelligence held in Puglia, at Borgo Egnazia, available at *www.vatican.va*, and also reproduced (among other places) in the appendix to the volume by V. PAGLIA, *L'algoritmo della vita. Etica e Intelligenza Artificiale*, Milano, 2024, p. 177 ff.: «Ed è ora lecito ipotizzare che il suo [dell'IA] uso influenzerà sempre di più il nostro modo di vivere, le nostre relazioni sociali e nel futuro perfino la maniera in cui concepiamo la nostra identità di essere umani». For an in-depth (and highly perceptive) philosophical-anthropological analysis of this technological reality in its relationship with human identity, see now *Nota* Antiqua et Nova *sul rapporto tra intelligenza artificiale e intelligenza umana*, a cura dei Dicasteri (della Santa Sede) per la Dottrina della fede e per la Cultura e l'Educazione, Città del Vaticano, 2025.

place human intelligence, operating independently of it and thereby bringing about a transfiguration of human identity [4]?

This radical and undoubtedly unsettling question seems, at least for now, to find a (temporarily) reassuring answer: there is, in fact, broad agreement – even among those who believe that such a scenario is not intended to remain unrealistic – in holding that, by now, AI is still not capable of letting itself free from the huge amount of data that is considered to be its fuel, data which – together with the algorithms that process it – continues to be provided by humans. In this respect, therefore, one can still speak of a technology that remains completely human [5].

2. The opportunities and risks linked to AI applications (a brief inventory)

And so, if it is a technology that remains under human control, it is essential to ensure that it serves humanity, promoting – rather than hindering – the full development of human freedom and dignity.

From this perspective, there is no doubt that the technologies making up the AI galaxy – which are, moreover, highly diverse from one another, thus raising equally diverse issues and suggesting the need to speak in the plural (artificial intelligences) [6] – might offer many positive opportunities and advantages that are so much important to be ignored [7]. Let's just think, for example, of the reduction of effort and fatigue in many areas of human activity, starting with physically demanding jobs; the important contribution AI can give in improving the efficiency of production processes; the innovations it may enable in agriculture, education, and culture; as well as the improvements in healthcare and the advancement of scientific research, thanks to its immense data-processing capabilities.

History teaches us, however – and it is enough here to recall, as a paradigmatic example in its immeasurable tragedy, the nuclear holocaust – that behind

[4] A. PATRONI GRIFFI, *op. cit.*, p. 11.

[5] A. PATRONI GRIFFI, *op. cit.*, p. 10 f.

[6] This also explains why, in scientific discourse, multiple definitions of the notion of AI recur.

[7] Thus, with specific reference to healthcare, P. BENANTI, *Intelligenza artificiale alleata della salute. Se le regole sono chiare*, in *Avvenire*, July 4, 2024 (supplement «È vita»), who writes: «L'intelligenza artificiale (IA) promette di rivoluzionare l'assistenza sanitaria, offrendo diagnosi più accurate, supporto personalizzato ai pazienti, scoperta di farmaci più rapida ed efficienza generale».

every scientific breakthrough lie many risks: the more serious they are, the greater the operational power of the breakthrough itself is.

Numerous and diverse, then, are the dangers lurking in the various and multifaceted applications of AI, so that it seems appropriate to draw up a brief inventory of the most significant among them [8].

It should be considered, to begin with, *digital surveillance*, which makes it possible to track the traces left by users in cyberspace and, as a result, to map and monitor individuals' mental and relational habits for commercial and political purposes, often without their knowledge, by shaping the flow of information in targeted and persuasive ways (for example think of the so-called *microtargeting*, which poses a serious threat to the integrity of electoral choices).

Subsequently, well-known dangers include targeted *disinformation* campaigns, which spread false information to manipulate public opinion, as well as the serious risks to *privacy* and *intellectual property* protection stemming from the vast collection and handling of data.

It should also be considered that algorithms can extract distorted information, reproducing and amplifying the injustices and biases present in the environments from which they originate. This phenomenon poses a serious threat to the *guarantees of impartiality* required by both public and private decision-makers and raises pressing ethical concerns, especially when the criteria used for decision-making are insufficiently transparent [9]. To grasp how impactful decisions left entirely or predominantly to the domain of algorithms can be, one need only to consider, for example, assessments of a person's creditworthiness when applying for a loan, suitability for a job or position, likelihood of recidivism for a convicted individual, eligibility for political asylum or social assistance, and more broadly, the adoption of *social credit scoring systems*, which inappropriately rank citizens according to opaque and questionable criteria.

Attention must also be drawn to the limitations placed on human freedom and responsibility in decision-making when choices are automatically determined by algorithmic calculations. This concern is particularly acute in contexts such as military command in wartime scenarios, the pervasive spread of sur-

[8] The inventory presented in the text is drawn, for the most part, from the Message *Intelligenza artificiale e pace* by His Holiness Pope Francis for the 57th World Day of Peace, January 1, 2024, available at *www.vatican.va*, and also reproduced (among other places) in the appendix to the volume by V. PAGLIA, *op. cit.*, p. 157 ff.

[9] Not to mention, furthermore, «che gli algoritmi disegnati per risolvere problemi molto complessi sono così sofisticati da rendere arduo agli stessi programmatori la comprensione esatta del come essi riescano a raggiungere i loro risultati»: thus spoke Pope Francis in the Address cited above in note 3.

veillance, the involvement of police forces in civilian life, as well as to the abyss of the so-called predictive justice entrusted to machines for the adjudication of legal disputes [10]. Not to mention the risks in the field of education, where algorithmic automatisms may weaken – and even extinguish – the capacity for developing *critical thinking*.

It is also important to highlight the significant social impact that AI applications may have on employment. In this regard, it has been observed that we are now facing the risk of mass unemployment among «colletti bianchi» – that is, those engaged in traditional middle-class professions – in a way reminiscent of the displacement of «colletti blu» during the earlier, less advanced phases of the various industrial revolutions [11]. A similar negative impact may also arise in terms of pollution and environmental sustainability, given that these technologies are highly *energy-intensive* and heavily reliant on *rare earth elements*, which are essential for the production of the hardware on which they depend.

Finally, apocalyptic scenarios are evoked with regard to the immensely powerful weapons systems that can be controlled by algorithms – systems whose destructive capabilities could potentially threaten the entire planet, should they escape human control.

3. The resulting need for ethical oversight of algorithms

The scenario just described clearly highlights the necessity, that has continually resurfaced throughout the history of civilization in response to successive scientific discoveries and their technological applications, to subject the various manifestations of AI (and, more broadly, of digital and information technologies) to a careful ethical scrutiny. This is all the more urgent given their formidable capacity to affect social relations and the existential sphere of individuals, as illustrated above. This need is so evident that it has given rise to a new branch of ethical reflection – which has by now been developing for some time –

[10] N. LIPARI, *Il diritto del nuovo millennio tra giurisdizionalizzazione ed algoritmo*, in *Acc.*, 2024, p. 15.

[11] P. BENANTI, *L'Intelligenza Artificiale sia al servizio del bene comune*, in *www.vaticannews.va*: «L'inizio della Rivoluzione industriale ha surrogato la forza muscolare; oggi vorremmo surrogare un po' quella che è la capacità cognitiva dell'uomo. (…) Se la prima Rivoluzione industriale ha impattato i colletti blu rendendoli meno necessari nel processo produttivo, l'Intelligenza Artificiale può impattare e impatterà tantissimo sui colletti bianchi, cioè su quei lavori che formano la classe media e, se non la gestiamo secondo dei criteri (…) anche di giustizia sociale, gli effetti potrebbero essere veramente devastanti o quanto meno molto forti nella capacità di coesione degli Stati democratici».

known as *algorethics*, a term derived from the algorithms themselves [12].

It is, as anyone can see, a matter of ensuring that the immense potential offered by these technologies is directed – to use fundamental moral (and spiritual) categories of meaning – toward the *good*, and not the *harm*, of humanity. That is, algorithms and the various applications of AI (as well as other digital and information technologies) should contribute to enhancing – not diminishing – the level of protection of human rights, individual freedom from the intrusion of public authorities and the arrogance of domestic and international economic and political powers; to promoting the defence of work and workers against all forms of degradation and exploitation; to protecting the most vulnerable and defenceless; to safeguarding the environment; and to advancing social justice.

In the final analysis, efforts must be made to ensure that these technologies become allies, not enemies, of humanity, understood according to the anthropological model embraced – returning now to the legal plane – by the Constitutions of the «dopo Auschwitz» era [13], in defense of its *inalienable dignity* [14]. That dignity is undoubtedly implicated – given the «nesso inscindibile» that mutually unites them – by the personalist principle enshrined in Article 2 of the Italian Constitution, which serves as the *Grundnorm* of the entire legal system [15].

[12] Now firmly embedded in university education – especially, as one would expect, in philosophy degree programs – and in postgraduate studies: see, for example, A. CARRARA, *Dall'algoretica alla roboetica: ecco le sfide*, in *Avvenire*, July 5, 2024, p. 26.

[13] I borrow the evocative expression – rich in symbolic resonance – from A. NICOLUSSI, *Un'ecologia giuridica per il futuro*, in *Vita pens.*, 1, 2012, p. 67; see also A. NICOLUSSI-C. RUSCONI, *Volti e risvolti della dignità umana. A settant'anni dall'art. 1 della Costituzione tedesca*, in *Jus*, 3, 2019, p. 4.

[14] Here stands out in all its moral pregnance the opening provision (Article 1) of the *Grundgesetz* of the Federal Republic of Germany, which solemnly proclaims: «Die Würde des Menschen ist unantastbar» («Human dignity is inviolable»).

[15] See, incisively, V. SCALISI, *L'ermeneutica della dignità*, Milano, 2018, p. 9 (where the A. discusses the «nesso inscindibile tra persona umana e dignità»), p. 22 («tra principio personalista e principio di dignità vi è nesso assai stretto, direi inscindibile, di reciproca e mutua implicazione») and *passim*. C. CAMARDI, *Sulla governance digitale europea: una proposta di confronto*, in *Acc.*, 2023, p. 8, reminds that even the *Declaration on the European Digital Decade: Digital Targets for 2030* – formalized in Decision (EU) 2022/2481 of the European Parliament and of the Council of 14 December 2022, establishing the Digital Decade Policy Programme 2030 – «fa perno sulla persona, e ribadisce perciò la primazia del principio personalistico dell'ordinamento giuridico europeo già sancito nella disciplina dei Trattati e della Carta» of Nice. In this same spirit, see now Recital 27 of the AI Act, Regulation (EU) 2024/1689 of the European Parliament and of the Council of 1 June 2024, laying down harmonised rules on artificial intelligence, published in the Official Journal of the European Union on 12 July 2024, which expressly emphasizes the need «che i sistemi di IA [siano] sviluppati e utilizzati come strumenti al servizio delle persone, nel rispetto della dignità umana e dell'autonomia personale (…)» (for an initial commentary on this regulation, see, among

No one can object that history never ceases to teach us that the achievements of science and technology, when perversely exploited in the pursuit of power and profit, risk unleashing destructive force upon the peaceful and orderly coexistence of peoples and nations, upon the preservation of nature and the environment, upon political and social freedoms, and – as mentioned above, and most importantly – upon the very idea of the inalienable dignity of every human being. Without this foundamental idea, the human rights arduously gained over the long course of our civilization's history would see their roots tragically severed, and thus be deprived of the vital lifeblood that sustains them.

And these enormous dangers of the perverse use of technology for purposes of political and/or economic domination – beyond being, in general terms, structurally inherent in the fusion of such technology with the vast machinery of capitalist power, which seeks to subordinate the entire organization of the world and the lives of its inhabitants to its own needs [16] – are, in the specific case of AI, amplified to an extraordinary degree by the fact that the entire system of such technologies is controlled by an oligopoly of nine global companies, none of which are European. These entities wield immense economic power [17] and – directly or indirectly, depending on the case – political power as well.

---

others, A. MANTELERO, *L'AI Act: la risposta del legislatore europeo alle sfide dell'intelligenza artificiale*, in *Acc.*, 2024, p. 191 ff.).

[16] The fusion between capitalism and technology (in Italian philosophical-legal literature, refer to N. IRTI-E. SEVERINO, *Dialogo su diritto e tecnica*, Roma-Bari, 2001) has by now become a fundamental topic in any serious reflection on the subject: within the vast literature, see most recently and notably, M. BARCELLONA, *Il diritto neoliberale dell'economia globalizzata e della società liquida*, in M. BARCELLONA-B. MONTANARI, *Potere e negoziazione. Il diritto al tempo del post-pensiero*, a cura e con prefazione di A. Lo Giudice, Roma, 2023, p. 30 ff., who evocatively describes the present era as «il tempo degli sponsali tra capitalismo e tecno-scienza». According to a well-known thesis, the «digital revolution» would have altered, in a very significant way, the characteristics of the mass production economic system, reshaping it into what has been called «Surveillance Capitalism»: S. ZUBOFF, *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*, New York, 2019 (trad. it. a cura di P. Bassotti, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Roma, 2019); the different term «Informational Capitalism» is used instead by J.E. COHEN, *Between Truth and Power. The Legal Constructions of Informational Capitalism*, Oxford, 2019, a term considered «più convincente» by G. RESTA, *Contratto e persona*, in *Tratt. contr.* Roppo, VI, *Interferenze*, 2ª ed., a cura di V. Roppo, Milano, 2022, p. 56, nota 13. This formulation has also been taken up by A. KAPCZYNSKI, *The Law of Informational Capitalism*, in 129 *Yale LJ*, 2020, p. 1460 ff.

[17] Consider that all the recalled nine multinational corporations have a market capitalization exceeding the amount of one trillion dollars. To grasp the scale, one need only note that the entire GDP of the United Kingdom amounts to 3.3 trillion dollars. This information is drawn from P. BENANTI, *L'Intelligenza Artificiale sia al servizio del bene comune*, cit.

4. The unavoidable subjection of digital technologies to the control of the existing legal system and the challenges of ensuring effective protection

This latter consideration clearly reveals, in the opinion of the writer, the groundlessness – and indeed the danger – of those ideological positions, which are not always innocent, that seek to exempt new digital technologies – and thus, ultimately, the economic giants behind them – from the rule of law [18].

Regardless of the creation of *ad hoc* regulatory frameworks (some of which already existing and others in varying stages of development), the first strategy to be pursued in order to counter this drift consists in bringing, as far as possible, the various technological applications under discussion back within the scope of existing legal norms and principles. This would allow us – at least to the extent permitted by their flexibility – to make use of the conceptual categories built around them. Indeed, the capacity of existing law to adapt to these emerging phenomena through interpretative evolution should by no means be underestimated [19].

In order to lend a degree of concreteness to the discussion at hand, it seems appropriate to propose a paradigmatic example which, from the perspective of civil law [20], may be identified here in the so-called *Smart Contract* [21].

---

[18] The opinion is widely shared: see, among others, *Alla ricerca del paradigma regolativo ottimale. La* governance *dell'I.A. tra legislazione e regolazione*, in *Tecn. dir.*, 2024, p. 279 ff.; K. FIORENZA, *Tecnologia e diritto. Una* odissea *negli spazi extracontrattuali*, Napoli, 2024, p. 24 f.

[19] On this point, reference may be made – in place of many others and also for further bibliographic leads – to the clear and persuasive pages of U. RUFFOLO-A. AMIDEI, *Diritto dell'Intelligenza Artificiale*, I, *Responsabilità. Contratto. Regolazione. Veicoli autonomi*, Roma, 2024, p. 69 ff.

[20] For an overall reflection on the relationship between AI and civil law, see most recently, P. PERLINGIERI, *Diritto civile e Intelligenza Artificiale*, in *Tecn. dir.*, 2025, p. 135 ff.

[21] Although *Smart Contracts* do not necessarily presuppose the use of AI-based technologies – on this point, see, among others, U. RUFFOLO-A. AMIDEI, *op. cit.*, p. 227 ff.; see also F. DI GIOVANNI, *Attività contrattuale e Intelligenza Artificiale*, in E. GABRIELLI-U. RUFFOLO (a cura di), *Intelligenza Artificiale e diritto*, in *Giur. it.*, 2019, p. 1677 ff. –, they nonetheless should be taken in consideration as a paradigmatic object of observation within a broader discourse on the relationship between law and digital technologies. It is well known that the debate surrounding the notion of *Smart Contract* (for an initial general overview of the concept, see G. SALITO, *Smart contract*, in *Dig. disc. priv.*, *Sez. civ.*, Agg. XII, Torino, 2019, p. 393 ff.) is highly animated and far from reaching a univocal definition (for a broader survey, see, among others, I. MARTONE, *Gli* Smart Contracts. *Fenomenologia e funzioni*, Napoli, 2022, *passim* e spec. pp. 53 ff. and 71-106). Without delving into the matter here, it suffices to note «che la posizione che pare preferibile è che non esiste una sorta di ontologia dello *smart contract*; gli *smart contract* sono protocolli di transazione che possono riguardare sia la fase di conclusione e di esecuzione di un contratto, che

The principal peculiarity of this figure lies in its self-executive nature [22], which stems from the automatic execution mechanisms it enables, due to its integration with distributed ledger technologies (DLT) [23] and, in particular, with the blockchain system [24].

This characteristic reveals a fundamental ambivalence: on one hand, it provides «certezza» (certainty) to contractual relations, ensuring with near-absolute reliability the actual performance of obligations (at least when the contract is executed entirely on-chain) [25]; on the other hand, however, it im-

solo quella di esecuzione di un contratto, così come quella di esecuzione di protocolli che nulla hanno a che fare con il contratto»: thus, among others, C. SCOGNAMIGLIO, *Lo* smart contract *in ambito bancario*, in M. PROTO (a cura di), *Intelligenza artificiale e rapporti bancari. Umano e non umano nelle relazioni tra intermediari e clienti*, Pisa, 2024, p. 35. This discussion refers specifically to the first two cases just described.

[22] The expression «autoliquidante» is used, among others, by A. D'ADDA, Smart contract *e diritto generale dei contratti*, in *Oss. dir. civ. comm.*, Speciale/2022, p. 105 ff.

[23] Mari. MAUGERI, Smart Contracts *e disciplina dei contratti*, Bologna, 2021, p. 20, testo e nota 6, rightly recalls that «spesso le nostre [dei giuristi] competenze sono insufficienti (…). A ciò si aggiunga che non abbiamo possibilità di verificare con facilità se la descrizione tecnica dei nuovi fenomeni corrisponda o meno a realtà». While keeping this *caveat* in mind, the clear attempt to frame the operating mechanisms of this type of technology made by I. MARTONE, *op. cit.*, p. 39 ff., nonetheless appears undoubtedly useful. This attempt includes abundant references to the literature and specific discussions on *Blockchain*; on the latter, see also, among others, N. TRAVIA, *La tecnologia Blockchain*, in E. BATTELLI (a cura di), *Diritto privato digitale*, 2ª ed. in collaborazione con G. D'Ippolito e T. Pellegrini, Torino, 2024, p. 329 ff.; G. PASCUZZI, *Il diritto dell'era digitale*, V ed., Bologna, 2020, p. 279 ff.; A. QUARTA-G. SMORTO, *Diritto privato dei mercati digitali*, Firenze, 2020, p. 22 ff.

[24] While one may concede the widely acknowledged point that the phenomenon of *Smart Contracts* does not necessarily coincide with DLT technologies (see, among others, and with specific reference to *Blockchain*, U. RUFFOLO-A. AMIDEI, *op. cit.*, p. 239), there is no doubt that these technologies – as I. MARTONE, *op. cit*., aptly observes – represent «a tutti gli effetti il "cuore pulsante" della contrattazione algoritmica» (p. 41), «il vero e proprio "trampolino di lancio" dello *Smart Contract* nella dinamica negoziale moderna» (p. 42). What holds true in general for DLTs is therefore also true, in particular, for Blockchain, which – as I. MARTONE, *op. cit*., notes – «rappresenta senz'altro la *species* più diffusa» (p. 42) of the former.

[25] A widely shared observation: see, among others, A. D'ADDA, *op. cit.*, p. 106; C. FERRARI, *Esecuzione del contratto e Intelligenza Artificiale*, in *De Iustitia*, 2024, p. 6; F. LONGOBUCCO, *Utopia di un'autonoma* Lex Criptographi(c)a *e responsabilità del giurista. Dallo* smart contract *al "*giusto rimedio*" del "*re-coding*" tra "*testo informatico*" e "*regolamento definitivo*"*, Napoli, 2023, p. 35 (work reviewed by L. COPPO, in *Tecn. dir.*, 2023, p. 469 ff.).

It has been observed, however, that the concrete possibility of non-performance or defective performance cannot be absolutely ruled out in cases involving the intervention of an oracle (for its notion, see below, note 56), particularly where malfunctions occur in its operation or it provides an incorrect assessment (P. CUCCURU, Blockchain *ed automazione contrattuale. Riflessioni sugli* smart contract, in *Nuova giur. civ. comm.*, 2017, II, p. 111). Such a possibility

plies a significant rigidity into this contractual model, precisely due to its inexorably automatic execution effects as a consequence of the inflexible logic embedded in the *if this, then that* sequence typical of computer code [26].

The model just described, while generating numerous and delicate legal issues [27], is nevertheless exalted by a radical ideological approach («cripto-anarchico») [28], which, leveraging the aforementioned automaticity of execution, tends to celebrate the "*Smart* paradigm" precisely for its alleged ability to legitimate itself *ex se* through the irresistible force of accomplished fact. This, in turn, would appear to place it ultimately beyond the reach of legal oversight [29], in a renewed (and enhanced) version of the *lex mercatoria*, reconfigured as *lex cryptographica* [30].

---

may also arise from the fact that the execution mechanism might fail to take into account the regulatory integrations pursuant to Article 1374 of the Italian Civil Code: in this sense, among many others, see, Mari. MAUGERI, *Contratto e nuove tecnologie*, in *Trattato del contratto* Roppo, VI, *Interferenze*, cit., p. 1166 ff.; non-performance may also occur in cases where the execution of a contract concluded *on-chain* involves the delivery of physical-world goods (E. BIVONA, Smart Contracts *e "interferenze" con la disciplina sui contratti: il sistema dei rimedi alla prova degli algoritmi*, in *Pers. merc.*, 2021, p. 795), due to the circumstance that «la pressoché assoluta garanzia dell'esecuzione del contratto (…) funziona solo in quanto le prestazioni riguardino oggetti digitalizzabili (come ad esempio criptovaluta)» (F. DI GIOVANNI, *op. cit.*, p. 1680).

[26] This sequence, upon closer examination, perfectly corresponds to the *if A, then B* scheme traditionally used by jurists to describe the logical structure of legal norms in general theory. For this notation, see M.F. CAMPAGNA, *Gli scambi attraverso algoritmi e il problema del linguaggio. Appunti minimi*, in *An. giur. ec.*, 2019, p. 154.

[27] For their analysis and the relevant bibliographic references, allow me to refer to V. VERDICCHIO, *Lo* smart contract *tra autonomia e autotutela*, in *Oss. dir. civ. comm.*, 2025, p. 3 ff.

[28] References in A. D'ADDA, *op. cit.*, p. 107; di «movimento cripto-anarchico *cypher-punk*» discusses Mari. MAUGERI, Smart Contracts *e disciplina dei contratti*, cit., pp. 25 (nota 9) and 55, who defines it as a «movimento socio-culturale (…) che teorizza la scelta della crittografia come modo per scardinare l'ordine esistente e avere una nuova democrazia»; on this point, see also, most recently, N. TRAVIA, *op. cit.*, p. 329 ff.

[29] In other, more evocative words, «Il *Νόμος* (…) si dovrebbe arrendere dinanzi alla *Τέχνη*»: thus, even though in a critical view, F. LONGOBUCCO, Smart Legal Contract *e "*giusto rimedio civile*" del* re-coding*:* Legal Protection by Design *e nuove declinazioni del* favor contractus, in *Jus civ.*, 2024, p. 767, who also speaks, again from the same perspective, of «dominio della tecnica sulla norma».

[30] F. LONGOBUCCO, *Utopia di un'autonoma* Lex Criptographi(c)a *e responsabilità del giurista*, cit., *passim*; E. BIVONA, *op. cit.*, p. 792 f. U.S. legal scholarship (L. LESSIG, *Code and Other Laws of Cyberspace*, New York, 1999) has expressed this line of thought through the now-famous formula *Code is Law*, which quickly spread worldwide and aims precisely to emphasize the (alleged) autonomy of computer code from the state legal system. In this regard, reference is also made to the concept of *lex informatica*: see, for example, M. GIULIANO, *La Blockchain e gli* Smart Contracts *nell'innovazione del diritto del terzo millennio*, in *Dir. inf.*, 2018, p. 995). For

From this point of view, with the advent of the *Smart Contract*, law would be imprisoned into the fact, and «la tecnica si fa[rebbe] contratto» – bearing in mind that, according to this view, we are not merely dealing with a contract executed through a new technical modality [31].

In this context, the elegant theoretical-sociological notion of *non-droit*, originally proposed some sixty years ago by Jean Carbonnier in his essay on "flexible law" [32], has been recalled and reinterpreted [33]. That is, a form of law that retreats to make room to the domain of facts, as they unfold in social reality – which, in this case, is understood as technological reality. In the phenomenon under discussion, to summarize, law would effectively disappear, since a *Smart Contract* could never give rise to issues of non-performance, not even in cases where the contract is affected by nullity.

As already noted, the *non-droit* approach is an ideological-epistemic perspective that, however, cannot be accepted. The legal system cannot remain indifferent to phenomena involving the circulation of wealth and exchanges – including those carried out through *Smart Contracts* – which, having general socio-economic implications and thus implicating, to some degree, the sphere of economic public order, cannot escape legal relevance and, consequently, to the reach of the law. This is all the more true when such transactions take place in sectors such as banking, finance, or energy, which are subject to intensive oversight by regulatory authorities that cannot be bypassed [34].

---

an overview, see, among many others, G. ALPA, Code is law*: il bilanciamento dei valori e il ruolo del diritto*, in *Contr. impr.*, 2021, p. 386 ff.

[31] A.M. BENEDETTI, *Contratto, algoritmi e diritto civile transnazionale: cinque questioni e due scenari*, in *Riv. dir. civ.*, 2021, p. 413; R. PARDOLESI-A. DAVOLA, *"Smart contract": lusinghe ed equivoci dell'innovazione purchessia*, in *Foro it.*, 2019, V, cc., p. 195 ff., spec. p. 195; C. SCOGNAMIGLIO, *op. cit.*, p. 34.

[32] J. CARBONNIER, *L'hypothèse du non-droit*, in *Arch. phil. dr.*, 1963, p. 55 ff. (cited by G.B. FERRI, *Il negozio giuridico tra libertà e norma*, V ed., Rimini, 1995, p. 80; ID., *Il diritto statuale e il suo* doppio, in *Quadr.*, 1993, p. 17); *amplius*, ID., *Flexible droit. Pour une sociologie du droit sans rigueur*, 10ª ed., Paris, 2001, (I ed., 1969; the VII French edition of 1992 was translated into Italian: *Flessibile diritto*, a cura e con un saggio di A. De Vita, Milano, 1997). For an in-depth reflection on the thought and works of the distinguished French jurist and sociologist, see the monograph by F.S. NISIO, *Jean Carbonnier*, Torino, 2002.

[33] F. DI CIOMMO, *Smart contract e (non-)diritto. Il caso dei mercati finanziari*, in *Nuovo dir. civ.*, 1-2019, pp. 257 ff., espec. p. 290 ff.

[34] Thus, among all, see C. SCOGNAMIGLIO, *op. cit.*, p. 39. On the use of *Smart Contracts* in the financial market and in the electricity sector, see Mari. MAUGERI, *Contratto e nuove tecnologie*, cit., respectively pp. 1174 ff. and 1182 ff.; For the insurance, banking, and financial markets, see also E. BATTELLI, *Le nuove frontiere dell'automatizzazione contrattuale tra codici algoritmici e* big data*: gli* smart contracts *in ambito assicurativo, bancario e finanziario*, in *Giust.*

It is therefore clear that the circulation of wealth via *Smart Contracts* cannot elude the grasp of the legal order [35]. Thus, even if the «autoliquidante» nature of such instruments precludes the use of traditional preventive remedies [36], and even if it is not possible to coercively interfere with the *Blockchain* – for example, to reverse undue transfers – this does not mean that there would be no reason to exclude remedies that were outside the blockchain environment [37].

From this perspective, remedial techniques that may still come into play – albeit only *ex post* and outside the blockchain environment – are quite evidently restitutory and/or compensatory, in response to economic operations that in some respect do not conform to legal standards [38]. Of course, this comes with all the well-known difficulties involved in actually securing of such forms of protection [39], since it is a universally shared knowledge that it is better not to make an undue performance in the first place than to attempt to recover it afterward or obtain compensation.

The example just discussed allows us to conclude that there is no reason to

civ., 2020, p. 681 ff.; G. GITTI, *Emissione e circolazione di criptoattività tra tipicità e atipicità nei nuovi mercati finanziari*, in *Banca borsa tit. cred.*, 2020, I, p. 13 ff.; also in ID., *Disciplina contrattuale del mercato e decisione robotica*, Brescia, 2020, p. 271 ff.

[35] Another crucial and sensitive issue – which cannot be explored in detail here – concerns the identification of the applicable legal system, since these contracts are often, so to speak, delocalized: see, among others, G. GITTI, *Tecnologie digitali, persona e istituzioni*, in *Riv. dir. civ.*, 2020, p. 1244 (also reproduced in ID., *Disciplina contrattuale del mercato e decisione robotica*, cit., p. 251), who discusses of «carattere "apolide" della tecnologia digitale e dell'algoritmo»; E. BIVONA, *op. cit.*, p. 792 ff.

[36] This refers to the dilatory exceptions provided for in Articles 1460-1462 of the Italian Civil Code.

[37] Mari. MAUGERI, Smart Contracts *e disciplina dei contratti*, cit., p. 52 f. Echoing these views, among others, are F. LONGOBUCCO, Smart Legal Contract *e "*giusto rimedio civile*" del* re-coding, cit., p. 770; L. REGAZZONI, *Alcune questioni in tema di* smart contracts *«assicurativi»: tra istanze di innovazione tecnologica e principi inderogabili dell'ordinamento giuridico*, in *Oss. dir. civ. comm.*, Speciale/2022, p. 133. For a general overview of smart contracts from a remedial perspective, see A. CINQUE, *Smart Contract: Legal Nature and Contractual Remedies*, *ivi*, p. 639 ff.

[38] C. SCOGNAMIGLIO, *op. cit.*, p. 37; Mari. MAUGERI, *op. ult. cit*., p. 53.

[39] The problems concerning the effectiveness of legal remedies (for a thorough general examination of the principle of effectiveness, see, among others, G. VETTORI, *Effettività fra legge e diritto*, Milano, 2020, as well as the reflections on the subject by V. VELLUZZI, *Effettività della tutela e giustificazione delle decisioni giudiziali. Riflessioni su* Effettività fra legge e diritto *di Giuseppe Vettori*, *in Riv. int. fil. dir*., 2024, p. 479 ff.) become even more pronounced in the case of *Smart Contracts* executed through decentralized DLTs, due to the anonymity these systems generally guarantee to the contracting parties. This anonymity also makes it difficult to verify whether the parties possess legal capacity: see, among others, Mari. MAUGERI, *op. ult. cit*., p. 56 ff.

totally exclude the application of existing law to new technologies, nor to deny its at least partial regulatory adequacy. At the same time, however, it reveals the limitations that the *jus conditum* may sometimes display – particularly in terms of the effectiveness of remedies – when tasked with governing digital and information technologies. This, in turn, calls for the development of new, specifically designed regulatory frameworks.

5. The underlying needs of digital technology regulation and the methods and characteristics of rulemaking

The reasons for rejecting the ideological approach that would seek to remove new technologies from the control of the legal system have already been identified.

Those reasons lead to the conclusion that the identified need for (new) regulation cannot be satisfactorily met by a system of rules, so to speak, «autoprodotte» by the operators of the sector themselves – such as codes of self-regulation and other forms of *soft law*. Even if one were to concede – drawing, for instance, on the Italian experience of the advertising self-regulation code [40] – that such an approach might offer some degree of utility [41], it nonetheless appears in all its evidence, in light of the enormity of the potential risks posed by these technologies, that it is in itself largely inadequate, and that ultimately it cannot dispense with a framework of (also) *heteronomous* regulation.

But how should we proceed in this direction? The limited scope of the present paper allows for no more than a few brief and general observations.

First and foremost, it must be acknowledged that – given the global relevance of the phenomenon under consideration – the ideal regulatory dimension should correspond to the spatial extent of the object to be regulated [42].

In this regard, although the well-known political, cultural, and interest-based divergences among States make it difficult – at least for now – to envision a truly global agreement, they have not prevented the recent conclusion of an important framework convention that establishes a clear set of principles aimed at ensuring

[40] It is sufficient, in this context, to refer – by way of brief mention – to G.F. CAMPOBASSO, *Diritto commerciale*, 1, *Diritto dell'impresa*, VII ed., a cura di M. Campobasso, Torino, 2013, p. 262 f.

[41] For some reflections on the matter, see, among others, U. RUFFOLO-A. AMIDEI, *op. cit.*, p. 51 ff.

[42] For some considerations, see G. FINOCCHIARO, *Intelligenza artificiale. Quali regole?*, Bologna, 2024, p. 51.

a balanced relationship between technological development, on the one hand, and the protection of human rights, democracy, and the rule of law, on the other [43]. This convention represents the first international treaty on AI; while it sets out important and binding reference standards, it will need to be implemented by the signatory States with a more or less broad margin of discretion.

In this direction – and necessarily narrowing the focus to domestic and EU law – the cornerstone of any regulation of digital and information technologies, starting with those associated with so-called AI, must be the respect for the *personalist principle* – briefly recalled above – enshrined in the Italian Constitution and, albeit with nuances not always perfectly aligned, also present in the founding Treaties of the European Union. As a consequence, the jurist – whether in the role of legal adviser to the Legislator or, more typically, as interpreter of the law – bears the responsibility of being guided by the North Star of human dignity and the principles of freedom and solidarity it entails.

The second key point to consider is that political and regulatory choices – while remaining faithful to the personalist principle and ensuring a sufficient set of legal guarantees – must also necessarily take into account that AI is, after all, a highly significant industrial sector, both economically and, even more so, strategically, with evident repercussions on the geopolitical stage. Accordingly, there is a need to construct a regulatory framework that enables – and, where possible, encourages – the strengthening of national and European business presence in this field, which, as previously noted, is dominated by global players from outside the European context (a Sino-North American oligopoly). It appears that this is precisely the path being pursued at the EU level by the so-called *AI Act*, which is based on a reasonable risk management approach – one that offers sufficient safeguards without deterring investment in the sector [44].

[43] This refers to the *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, signed on 5 September 2024 by Andorra, Georgia, Iceland, Norway, Moldova, San Marino, the United Kingdom, Israel, the United States of America, and the European Union. The Convention imposes on the signatory States the obligation to regulate the entire life cycle of artificial intelligence systems, with particular emphasis on *transparency*, *accountability*, and *privacy protection*.

[44] See, among others, A. MANTELERO, *op. cit.*, p. 196, who observes that «l'approccio adottato dal legislatore europeo si connoti per un notevole pragmatismo, guardando ad una regolamentazione incentrata sui rimedi *ex ante*, in chiave di gestione del rischio ed approccio *by-design* al prodotto, piuttosto che sui rimedi risarcitori *ex post* (…). Questo perché il rimedio tradizionale della responsabilità civile mal si concilia con un contesto connotato da complessità tecnologica, operatori globali con grandi capacità finanziarie, polverizzazione dei danni e, nell'ottica di stimolare la fiducia nelle nuove tecnologie (la c.d. *trustworthy AI*), necessità di garantire ambienti tecnologici sicuri piuttosto che rimedi in caso di disastrose conseguenze». The A. further emphasizes the need to provide residual rules on tort law.

Having said that, it is often asserted that regulation of the technological sector in question should be framed in terms of principles rather than rules [45]. The reasoning behind this claim is that the pace of ordinary legislative processes – carried out through political assemblies – is incompatible with the rapid and continuous evolution of these technologies. As a result, rules produced in this way run the risk of becoming quickly obsolete, if not already outdated at the moment of their enactment, and in any case of arriving consistently too late [46].

It must be observed, on the other hand, that while a good regulation cannot stand without clear guiding principles – principles to which, of course, should also be given proper weight at the interpretative and applicative levels – it likewise cannot dispense with a set of more detailed rules. Such rules serve to give (so to speak) shape and measure to the principles to be implemented, while also governing their mutual interactions (and possible collisions), thereby containing within acceptable limits the practical uncertainties inevitably generated by their intrinsic structural indeterminacy – which, at times, may even verge on vagueness [47].

In their absence, such principles would inevitably be interpreted and applied differently by each interpreter and – even before that, and more importantly – by each addressee of the regulation [48]. This would undermine the unavoidable

[45] This refers primarily to the principles set out in the European Union's 2019 document *Ethics Guidelines for Trustworthy AI*, which also clearly inspired the *AI Act*. These include: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being; accountability (see, among others, G. FINOCCHIARO, *op. cit*., p. 47). These parameters appear to be grounded in the broader precautionary principle.

[46] These widely shared and persuasive observations (see, among others, M. ANGELONE, *op. cit*., p. 287) clearly show that, although detailed regulation is indeed necessary, it cannot be entrusted solely to the ordinary legislative activity of elected parliamentary chambers (see *infra* in the text).

[47] If it is largely undisputed that every legal provision – even when formulated through a regulatory legislative technique (i.e., based on sufficiently detailed rules) – entails a certain degree of indeterminacy and vagueness, it appears equally evident, however, that these features are present to an even greater extent in a regulatory framework based exclusively on principles. While such principles are indispensable for providing both the axiological and rational foundation of the rules – and for enabling a sufficiently flexible interpretation during their application (a role also fulfilled, from another angle, by general clauses) – they are not sufficient on their own: «L'assenza di una griglia di regole più minute consegna l'ordinamento all'arbitrio individuale, che forse (…) è un male peggiore dell'eccesso di regolamentazione»; thus, ultimately, everything depends on finding the right balance between overarching principles and detailed rules (G. TERRANOVA, *Il diritto e il suo linguaggio. Metodi, teorie, parole del diritto commerciale*, Milano, 2020, p. 55).

[48] Illustrative in its clarity is the example – recalled by G. TERRANOVA, *op. cit*., p. 187 (and drawn from F. SCHAUER, *Di ogni erba un fascio. Generalizzazioni, profili, stereotipi nel mondo della giustizia*, Bologna, 2003) – offered by a concrete North American experience: «In una certa fase storica, nel Montana si pensò di sostituire ai limiti di velocità per il traffico stradale, indicati

requirement of at least a tendential *certainty* of the law – which is itself, it should be noted, a fundamental (value and) normative principle. This requirement is all the more necessary in the field at hand in order to implement that by design form of preventive regulation which, as is well known, is essential to legally shape digital environments from the outset. And such regulation can hardly do without the presence of (at least a minimal number of) clear and detailed rules [49], particularly of a *technical nature* [50].

That being established, the aforementioned criticisms of detailed regulation enacted through legislative processes nonetheless prove to be far from unfounded. In addition to what has already been noted regarding the slow and often inadequate pace of parliamentary response, it must also be considered that, «in campi o materie caratterizzati da una spiccata spinta evolutiva, i procedimenti normativi scontano forti *deficit* informativi e di competenze/conoscenze, necessitando per contro di informazioni dettagliate, granulari, attendibili che sono sovente inaccessibili ovvero nell'esclusiva disponibilità degli stessi destinatari della regolazione» [51].

These considerations support the need to entrust the governance of AI to one or more independent authorities, endowed with the classic prerogatives – regulatory, administrative, supervisory, adjudicatory, and sanctioning powers (whether monetary or otherwise) – that such bodies typically exercise in so-called «settori sensibili» (*domaines sensibles*) placed under their oversight [52].

All of this must, of course, be framed within a clear structure of guiding principles and boundaries, which – and here rightly so – must be established by ordinary legislation, in keeping with the political responsibility of Parliament [53].

---

in appositi cartelli, un principio generale, che si limitava a imporre una guida prudente e ordinata. L'esito dell'esperimento fu disastroso, giacché ogni conducente interpretava a modo suo i predetti limiti prudenziali, con la conseguenza che si moltiplicarono a dismisura gli incidenti automobilistici e i morti su strada. A causa di questa tragica esperienza si tornò alla disciplina preesistente».

[49] Well known is the observation that the digital sphere is governed «soprattutto [da] "forme" architettoniche. (…) il dispositivo principe di governo dello spazio digitale è il codice informatico che può erigere steccati, impedendo e/o facilitando l'accesso soltanto a determinate condizioni; in altri termini è lo strumento per eccellenza in grado di governare la rete e orientare i valori cui si ispira» (thus, among all, K. FIORENZA, *op. cit.*, p. 27 f.). It follows that, if the law wishes to influence digital reality, it must do so primarily by design – that is, by regulating the way in which digital ecosystems are to be built through the computer codes that shape them. And it seems clear that for this to be achieved effectively, principles and general clauses – though absolutely necessary – are not sufficient on their own; specific technical rules are also required.

[50] Among all, U. RUFFOLO-A. AMIDEI, *op. cit.*, p. 70.

[51] M. ANGELONE, *op. cit.*, p. 287.

[52] M. ANGELONE, *op. cit.*, p. 287 f.

[53] See, among others, once again, M. ANGELONE, *op. cit.*, p. 287.

6. Legal scholars and legal science facing the digital revolution: towards the definitive overcoming of the «two cultures»

The reflections developed so far prompt a final consideration, this time concerning the impact that new digital and information technologies are having – or, in any case, ought to have – on the role of the jurist and on the method of legal science.

To explore this aspect more deeply, it seems useful to return once again to the example of the *Smart Contract* [54].

The inescapable «autoliquidante» mechanism previously discussed places significant strain on the systemic devices of invalidity and regulatory integration under Article 1374 of the Italian Civil Code. On the one hand, it results in the automatic execution of the agreement as encoded by the parties in the software code – even in cases where the agreement is unlawful or otherwise null and void. On the other hand, such execution does not take into account regulatory integrations, not even those deemed mandatory, as provided by the combined provisions of Articles 1339 and 1419(2) of the Civil Code, which introduce heteronomous norms into the content of the contract.

Furthermore, the aforementioned «autoliquidante» nature of the Smart Contract tends, in practice, to neutralize the legal relevance of supervening events, since its automatic execution is generally incapable of adapting to changes in the original factual circumstances. This characteristic also disrupts the overall structure of the ordinary system of judicial and extra-judicial remedies: on the one hand, there is an inherent and almost ontological difficulty in raising dilatory exceptions under Articles 1460-1462 of the Civil Code; on the other hand, there is a tangible risk that the weaker party may be exposed to an excessive – so to speak – self-help advantage wielded by the stronger party.

The negative consequences identified – stemming from the immutability, irreversibility, and unstoppable nature of *Smart Contracts* – ultimately risk causing a significant impairment of the protections established by the legal system, as well as a disruption of the underlying balance of interests.

It is therefore understandable that the European Parliament has expressed concern and has emphatically urged the Commission to provide, in relation to such contracts, for mechanisms capable of *halting* and *reversing* their execution and the related payments [55]. Similarly, the recent Regulation (EU) 2023/2854

[54] For the in-depth legal analyses of the *Smart Contract* issues explored immediately after in the text, reference is made to V. VERDICCHIO, *op. cit.*, p. 3 ff.

[55] European Parliament Resolution of 20 October 2020, containing recommendations to the Commission on the Digital Services Act, in *G.U.U.E.*, 6 October 2021, C-404, pp. 31 ff., esp.

(the so-called *Data Act*) has identified, as a necessary requirement for smart contracts, the provision of «[…] b) safe termination and interruption, to ensure that a mechanism exists to terminate the continued execution of transactions», estabilishing «that the smart contract includes internal functions which can reset or instruct the contract to stop or interrupt the operation, in particular to avoid future accidental executions (…)» [art. 36 (1)].

Indeed, the technical feasibility of incorporating functions capable of halting execution and modifying the terms of *Smart Contracts* has been the subject of reflection for some time. There have even been proposals – admittedly ambitious – to assign the function of oracle [56] to a state judge or (less irrealistically) to a private arbitrator, thereby enabling them to manage the critical issues associated with this contractual model [57].

It is not possible here to dwell on a detailed description of the proposed technical solutions [58], but it is clear that the design and potential implementation of

37 (point no. 35). Mari. MAUGERI, Smart Contracts *e disciplina dei contratti*, cit., p. 64, considers this request too intrusive, as it is «distonic[a] rispetto al modello di funzionamento dello Smart Contract», which is characterized by speed and efficiency.

[56] On the notion of the oracle, which constitutes the only means of linking external reality to the otherwise entirely self-referential technological and informational reality, see, among others, I. MARTONE, *op. cit.*, p. 60; as well as Mari. MAUGERI, *op. ult. cit*., p. 34 f.: «Si definisce oracolo la macchina o l'umano che manda l'*input* di ricezione alla macchina. L'oracolo è, dunque, un'entità terza che manda l'*input* al sistema. Si distingue fra oracolo automatico, ad esempio un sensore connesso con il *web site*, e oracolo uomo, si pensi al corriere che ha consegnato il pacco o a un esperto con poteri di valutazione. Axa, ad esempio, ha immaginato un contratto con il quale in caso di ritardo dell'aereo si avrà automaticamente il pagamento di una somma. In questo caso l'oracolo sarà il sistema di comunicazione degli orari di arrivo degli aerei, che verrà consultato e determinerà l'immediato trasferimento della somma in caso di ritardo». On the subject, see also M. EPIFANIA, *Questioni aperte in tema di* Smart Contract *e* Blockchain*: il «problema dell'oracolo»*, in *Corti salern.*, 2022, p. 621 ff.

[57] With regard to the intervention of third-party institutions, it has been noted, moreover, that there is a «contrasto con la natura che si vorrebbe tendenzialmente disintermediata delle piattaforme DLT, luogo di naturale sviluppo dello *smart contract*»: thus, among others,, A. D'ADDA, *op. cit.*, p. 113.

[58] For the identification of which, reference is made – including for the necessary bibliographic insights – to the following recent studies: C. ATTANASIO, *Inadempimento dello* smart contract*, sistema rimediale e tutela effettiva*, in *Riv. dir. civ.*, 2024, p. 732 ff.; T. PELLEGRINI, *Gli* smart contract, in E. BATTELLI (a cura di), *Diritto privato digitale*, cit., p. 371 ff.; F. LONGOBUCCO, Smart Legal Contract *e "*giusto rimedio civile*" del* re-coding, cit., pp. 762 ff., espec. 768 ff. This A. – following the proposal advanced by M. ZALNERIUTE-L. BENNETT MOSES-G. WILLIAMS, *The Rule of Law 'By Design'*, in *Tulane LR*, 95 (5), 2021, p. 1063 ff. – argues that should be «incoraggiati tutti i fenomeni incentrati sulla *Rule of Law by design*»; see also ID., *Utopia di un'autonoma* Lex Criptographi(c)a *e responsabilità del giurista*, cit., p. 102 ff.; to manage disputes relating to smart contracting E. BIVONA, *op. cit.*, p. 796 f., suggests adapting *Online Dispute Resolution (ODR) systems* to the blockchain environment; for an initial general overview of these systems, see, among others, E. MINERVINI (a cura di), *Le* online dispute resolution (ODR), Napoli, 2016.

technological mechanisms capable of adequately addressing the complex legal issues raised by smart contracting cannot proceed without a joint effort – and thus an ongoing dialogue – between programmers (IT specialists and other tech professionals involved) and jurists [59]. Such collaboration is essential in order to develop technological artefacts capable of operating in harmony with the rules and principles of the legal system [60].

This reflection can quite evidently be extended beyond the specific domain of *Smart Contracts* to encompass the entire scope of new technologies – first and foremost, those that embody the complex and multifaceted phenomenon of so-called Artificial Intelligence.

This new pervasive reality calls for a shift – indeed, an epistemological paradigm shift – on the part of jurists and legal science; and the same, conversely, may be said of computer scientists and other scholars in the «hard sciences» who are engaged in this technological revolution [61].

---

[59] F. LONGOBUCCO, Smart Legal Contract *e "*giusto rimedio civile*" del* re-coding, cit., p. 771 (where further bibliographic references can be found) calls for the specialization of the jurist as a *Legal Engineer*, capable of engaging in a «dialogo costante con lo sviluppatore che è chiamato a "disegnare" *ab origine* una architettura informatica» compatible with legal rules and principles, in line with the *Rule of Law by design approach* (see the immediately preceding note); on the evolution of legal professions in relation to the digital reality, see, in greater detail, ID., *Utopia di un'autonoma* Lex Criptographi(c)a *e responsabilità del giurista*, cit., p. 35 ff.; see also what is noted in the subsequent footnote 68.

[60] The insightful perspective of M. PROTO, *Funzioni delle 'macchine intelligenti' e compiti dell'interprete*, in ID. (a cura di), *Intelligenza artificiale e rapporti bancari*, cit., p. 11, highlights, moreover, that law itself also represents, so to speak, an artificial technique, «che faticosamente costituisce regole osservando la realtà e tipizzandola, cioè convertendola in norme generali e astratte»; thus making it, in this sense, comparable to technological reality.

[61] It is reassuring to note that the aforementioned need for discussion and dialogue is also well recognized by scholars of computer science; emblematic, in this regard, are the words spoken by G. CANFORA during the *Institutional Greetings* at the Conference «Piattaforme online, dati e intelligenza artificiale tra interessi pubblici e garanzie dei privati», held in Benevento on June 13-14, 2024, at the University of Sannio (where he serves as Rector), as reported in *these Review*, 2025, p. 135 ff.: «Si sente spesso dire che abbiamo bisogno di più ingegneri, ed è indubbiamente vero! Ma abbiamo bisogno che i nostri ingegneri sviluppino un'attitudine umanista, dobbiamo dare loro gli strumenti concettuali per maturare piena consapevolezza delle conseguenze delle scelte tecnico-scientifiche sulla società e sulle vite delle persone. Al contempo, abbiamo bisogno che chi si occupa di discipline non tecniche, gli avvocati, i filosofi, i letterati, i sociologi, gli economisti, per fare qualche esempio, conosca più da vicino le tecnologie digitali, per comprendere fino in fondo la portata dei cambiamenti che emergono dall'interazione fra le tecnologie e dai modi in cui queste vengono utilizzate». Naturally, this need is increasingly recognized by legal scholars as well: for an initial reference, see, among others, E. MINERVINI, *Note brevi sull'intelligenza artificiale*, *ivi*, p. 195; F. LONGOBUCCO, Smart Legal Contract *e "*giusto rimedio civile*" del* re-coding, cit., p. 775; G. GITTI, *Presentazione*, in *Oss. dir. civ. comm.*, Speciale/2022, p. 5 ff.

From the perspective of jurists – both theorists and practitioners – this implies the need to deepen their understanding of the *technological domain* that law aspires to regulate [62]. To be clear, the era of purely doctrinal legal theories are long gone. For quite some time now, jurists have been fully aware that theirs is a practical science, and that in order to perform their task effectively, they must engage in dialogue with other disciplines that enable them to comprehend the social environment subject to legal regulation. Consequently, for just as long, jurists have been in conversation with philosophy, history, sociology, economics, political science, and more broadly, with the other social sciences [63].

What is now required is a step forward – the dismantling of the wall of separation (and mutual incomprehension) between scientific and humanistic culture, a divide famously denounced by C.P. Snow as early as 1959 in his seminal book *The Two Cultures* [64]. Jurists must jump over the fence – a task they have already begun to undertake, for instance, by engaging with genetics, biology, and medicine in its various branches when addressing bio-legal questions relating to the beginning and end of life. What is now needed is the recognition that «[p]er comprendere davvero i temi posti dalle nuove tecnologie, ancor più dall'intelligenza artificiale, e trovare risposte e regole, occorre aprirsi alle conoscenze dei diversi punti di vista e prospettive della cultura scientifica» [65], allowing it to engage in dialogue with law and the other social sciences. What is ultimately required is an awareness of the unity of the «due culture» [66].

[62] «La conoscenza del fenomeno è quindi pregiudiziale ad ogni valutazione [*sub specie juris*] dello stesso»: thus E. GABRIELLI-U. RUFFOLO, *Introduzione*, in ID. (a cura di), *Intelligenza Artificiale e diritto*, cit., p. 1657. The programmatic manifesto of such a cultural commitment could be encapsulated in the evocative formula «forma giuridica e materia tecnologica», reminiscent – *mutatis mutandis* – of the title of a famous essay by L. MENGONI, *Forma giuridica e materia economica*, in *Studi in onore di Alberto Asquini*, III, Padova, 1963, p. 1075 ff.; e in *Jus*, 1966, p. 5 ff.; and, more recently, in ID., *Diritto e valori*, Bologna, 1985, p. 147 ff.

[63] To cite just one example, A. QUADRIO CURZIO, *Presentazione*, in A. CARLEO (a cura di), *Calcolabilità giuridica*, Bologna, 2017, p. 7, speaks of «antica e importante complementarità tra le scienze giuridiche e le scienze economiche»; however, on closer examination, there are those who, on this very point, instead denounce «la disgiunzione tra il sapere tecnico-economico e la conoscenza giuridica» as «uno dei problemi più complessi e discussi del nostro tempo» (thus G. LEGNINI, *Introduzione*, in the aforementioned volume, p. 11). For a comprehensive examination of the relationship between law and economics, most recently, G. AZZARITI-S. BAGNI-M. CARDUCCI-A. SOMMA (a cura di), *Mercato istituzioni e regole. Sul difficile dialogo tra diritto ed economia*, Roma, 2025.

[64] C.P. SNOW, *The Two Cultures and the Scientific Revolution*, Cambridge UP, 1959.

[65] A. PATRONI GRIFFI, *op. cit.*, p. 12.

[66] For an overview, most recently, T. BONGO-G. CAPASSO-A. EREDITATO-M. FARISCO-O. SAMPIETRO (a cura di), *L'unità delle due culture. Studi offerti a Ortensio Zecchino per i suoi ottant'anni*, I-II, Soveria Mannelli, 2023. And see also J. BROCKMAN, *The Third Culture*, New York, 1995 (trad. it. di L. Carra, *La terza cultura*, Milano, 1995).

The stakes are exceedingly high. For the dialogue and mutual understanding just invoked – beyond fulfilling the important gneosological and methodological need just highlighted – is likely also essential to prevent the «marginalizzazione totale» of the specific knowledge tradition embodied in legal science [67]. If it fails to rise to this challenge, legal science risks being overwhelmed by the expansive rise of modern technologies and the diverse bodies of knowledge cultivated by the techno-sciences that underpin and fuel them [68].

[67] Thus Mari. MAUGERI, *Contratto e nuove tecnologie*, cit., p. 1150, to whom the expression quoted in the text is attributed. Although particularly serious in the present time due to the noted powerful rise of techno-sciences, the issue of the marginalization of legal knowledge «nella più generale storia della cultura di un popolo» – and, consequently, of the role of jurists and of legal science «nella comune considerazione» – is, however, a very recurrent concern in legal literature and far more longstanding, to the point of arguably representing an authentic τόπος (tòpos). See, by way of example, G.B. FERRI, *Il diritto statuale e il suo* doppio, cit., p. 21 (to whom the preceding quotations belong). Equally longstanding, moreover, are the attempts made by the most forward-looking members of the legal community to redeem this branch of knowledge from isolation or, at the very least, from subordination to broader cultural trends, through the deliberate implementation of initiatives inspired by methodological openness grounded in multi- and transdisciplinary dialogue. Among the many possible examples, it is worth recalling here the cultural project that took shape with the publication, beginning in 1972, of the *Quaderni fiorentini per la storia del pensiero giuridico moderno*, published by Giuffrè: as stated in the *Pagina introduttiva* of the first 1972 volume, always signed by Paolo Grossi (now reproduced in ID., *Trent'anni di pagine introduttive. Quaderni fiorentini 1972-2001*, Milano, 2009, p. 3 ff., from which this passage is quoted), one of the guiding motivations of that successful and commendable editorial undertaking was precisely the aspiration to take «un primo passo nella strada (…) che ha, al suo sbocco finale, la sottrazione della scienza giuridica dai margini della cultura nei quali è oggi relegata»; see also the preface – significantly entitled *I segni dei tempi* – by P. CAPPELLINI to the above-mentioned volume, which, in recalling the founding journey of the Quaderni, reports (p. XIX ff.) on a letter sent to Paolo Grossi by his first student, Mario Sbriccoli, who underscored the necessity «del recupero della scienza giuridica alla cultura», while also seeking to identify the underlying reasons for the «perifericità» of that body of knowledge.

In Italian-language literature, is now available, in the form of a comprehensive treatise (*Trattato e-Agorà*), an in-depth interdisciplinary study – primarily devoted to digital platforms – which jointly examines the technological, philosophical, and legal aspects of the issue, with the latter articulated across the various branches of legal scholarship: R. BOCCHINI (a cura di), *Le piattaforme digitali*, Torino, 2025.

[68] This necessity must, of course, be reflected in the educational system, primarily at the university level, which, however, has already begun to move in this direction and continues to do so in our country. There are now several examples of law degree programs in Italy that include specific tracks – under various designations – aimed at providing an in-depth study of the relationship between law and technology.

AI LAW

INTERNATIONAL REVIEW
OF ARTIFICIAL INTELLIGENCE LAW

El impacto de la inteligencia artificial en la producción científica *

Fabio Addis

Professore ordinario di Diritto privato
nell'Università di Roma "La Sapienza"

Abstract

(EN): The article examines the impact of generative artificial intelligence on scientific production, framing it as an epistemological discontinuity. The author introduces the principle of "epistemic self-responsibility," which requires researchers to take conscious responsibility for the tools they employ. The paper identifies the essential criteria for the responsible use of AI in research – conditional legitimacy, transparency, human oversight, exclusion of authorship, and restrictions in peer review – and highlights the decisive role of law, academic institutions, and scientific journals in shaping appropriate rules and policies. AI is presented not as a threat but as an opportunity for methodological renewal, enhancing processes without replacing human creativity and responsibility. The article concludes with a personal reflection, invoking three ethics – prudence, trust, and care – as the conditions for a fruitful balance between technology and science.

(ES): El artículo examina el impacto de la inteligencia artificial generativa en la producción científica, enmarcándolo como una discontinuidad epistemológica. El autor introduce el principio de "autorresponsabilidad epistémica", que exige a los investigadores asumir de manera consciente la responsabilidad por las herramientas que emplean. El trabajo identifica los criterios esenciales para el uso responsable de la IA en la investigación – legitimidad condicionada, transparencia, supervisión humana, exclusión de la autoría y limitaciones en la *peer review* – y destaca el papel decisivo del Derecho, de las instituciones académicas y de las revistas científicas en la configuración de normas y políticas adecuadas. La inteligencia artificial se presenta no como una amenaza, sino como una oportunidad de renovación metodológica, capaz de mejorar los procesos sin sustituir la creatividad y la responsabilidad humanas. El artículo concluye con una reflexión personal, invocando tres éticas – prudencia, confianza y cuidado – como condiciones para alcanzar un equilibrio fecundo entre tecnología y ciencia.

(FR): L'article examine l'impact de l'intelligence artificielle générative sur la production scienti-

* Enriquecido con las referencias bibliográficas que constituyen su principal fuente de inspiración, el presente texto es la ponencia titulada «Impacto de la Inteligencia Artificial sobre la producción científica», presentada en el III Congreso Internacional sobre *Derecho e Inteligencia Artificial. Su proyección en los derechos fundamentales*, celebrado en Bilbao los días 27 y 28 de marzo de 2025 en la Facultad de Derecho de la Universidad de Deusto.

fique, en l'envisageant comme une discontinuité épistémologique. L'auteur introduit le principe de «responsabilité épistémique personnelle», qui exige des chercheurs qu'ils assument consciemment la responsabilité des outils qu'ils emploient. Le texte identifie les critères essentiels d'un usage responsable de l'IA dans la recherche – légitimité conditionnelle, transparence, supervision humaine, exclusion de l'autorat et limitations dans l'évaluation par les pairs – et souligne le rôle décisif du droit, des institutions académiques et des revues scientifiques dans la définition de règles et de politiques appropriées. L'intelligence artificielle est présentée non comme une menace, mais comme une occasion de renouveau méthodologique, capable d'améliorer les processus sans se substituer à la créativité et à la responsabilité humaines. L'article se conclut par une réflexion personnelle, invoquant trois éthiques – prudence, confiance et soin – comme conditions d'un équilibre fécond entre technologie et science.

(DE): Der Artikel untersucht die Auswirkungen generativer künstlicher Intelligenz auf die wissenschaftliche Produktion und versteht sie als eine epistemologische Diskontinuität. Der Autor führt das Prinzip der „epistemischen Selbstverantwortung" ein, das von Forschenden verlangt, eine bewusste Verantwortung für die von ihnen verwendeten Werkzeuge zu übernehmen. Der Beitrag identifiziert die wesentlichen Kriterien für den verantwortungsvollen Einsatz von KI in der Forschung – bedingte Legitimität, Transparenz, menschliche Aufsicht, Ausschluss der Autorschaft und Einschränkungen im *Peer Review* – und hebt die entscheidende Rolle des Rechts, der akademischen Institutionen und der wissenschaftlichen Zeitschriften bei der Ausgestaltung geeigneter Regeln und Politiken hervor. Künstliche Intelligenz wird nicht als Bedrohung, sondern als Chance für eine methodologische Erneuerung dargestellt, die Prozesse verbessern kann, ohne menschliche Kreativität und Verantwortung zu ersetzen. Der Artikel schließt mit einer persönlichen Reflexion, in der drei Ethiken – Vorsicht, Vertrauen und Fürsorge – als Voraussetzungen für ein fruchtbares Gleichgewicht zwischen Technologie und Wissenschaft beschworen werden.

(PT): O artigo examina o impacto da inteligência artificial generativa na produção científica, enquadrando-o como uma descontinuidade epistemológica. O autor introduz o princípio da "autorresponsabilidade epistémica", que exige que os pesquisadores assumam, de forma consciente, a responsabilidade pelas ferramentas que utilizam. O texto identifica os critérios essenciais para o uso responsável da IA na pesquisa – legitimidade condicionada, transparência, supervisão humana, exclusão de autoria e restrições na *peer review* – e destaca o papel decisivo do Direito, das instituições acadêmicas e das revistas científicas na formulação de regras e políticas adequadas. A inteligência artificial é apresentada não como uma ameaça, mas como uma oportunidade de renovação metodológica, capaz de aprimorar os processos sem substituir a criatividade e a responsabilidade humanas. O artigo conclui com uma reflexão pessoal, evocando três éticas – prudência, confiança e cuidado – como condições para um equilíbrio frutífero entre tecnologia e ciência.

(IT): L'articolo analizza l'impatto dell'intelligenza artificiale generativa sulla produzione scientifica, evidenziandone il carattere di discontinuità epistemologica. L'autore propone un approccio fondato sul principio di "autoresponsabilità epistemica", che richiede al ricercatore di assumere consapevolmente la responsabilità degli strumenti utilizzati. Vengono individuati i criteri minimi di un uso responsabile dell'IA nella ricerca – legittimità condizionata, trasparenza, controllo umano, esclusione dell'autorialità e limitazioni nella *peer review* – e si sottolinea il ruolo decisivo del diritto, delle istituzioni accademiche e delle riviste scientifiche nell'elaborare regole e policy adeguate. L'IA è interpretata non come minaccia, ma come occasione di rigenerazione metodologica, capace di potenziare i processi senza sostituire la creatività e la responsabilità umana. L'articolo si conclude con una riflessione personale, che richiama le tre etiche della prudenza, della fiducia e della cura, delineando le condizioni per un equilibrio fecondo tra tecnica e scienza.

Sumario: 1. El contexto del cambio. – 2. Ciencia, responsabilidad y técnica: el principio de autorresponsabilidad. – 3. Uso responsable de la IA: una gramática común. – 4. Hacia una nueva metodología de la investigación. – 5. El papel del Derecho y de las instituciones académicas. – 6. Una voz personal: ética de la prudencia, ética de la confianza, ética del cuidado. – 7. La IA en la ciencia: una oportunidad de regeneración.

1. El contexto del cambio

La irrupción de la inteligencia artificial generativa en los procesos cognitivos, creativos, comunicativos – y, hoy en día, también científicos – no constituye un simple acontecimiento técnico.

Es, en todos los sentidos, una discontinuidad epistemológica [1].

El hecho de que programas como ChatGPT, Claude, Gemini, Copilot, o herramientas como Elicit, Scite y Writefull sean capaces de generar textos estructurados, sugerir bibliografías coherentes o sintetizar fuentes complejas, impone una revisión de los presupuestos sobre los que se funda el propio acto de "investigar".

Pero, lo que hoy se cuestiona – se me permita decirlo con claridad – no es el valor intrínseco de la tecnología. No es su legitimidad. Ni siquiera su eficacia.

Lo que está verdaderamente en juego es la relación entre saber y responsabilidad, entre el poder de producir y el deber de rendir cuentas [2].

Y aquí el Derecho tiene una tarea noble: no la de contener la tecnología por temor, sino la de acompañarla con lucidez [3]. Regular sin paralizar. Discernir sin demonizar. Abrir sin ceder [4].

---

[1] T. KUHN, *The Structure of Scientific Revolutions*, Chicago, 1962, p. 92: «A paradigm shift does not simply involve a change in scientific theory, but a transformation in the rules by which scientific practice itself is conducted».

[2] «Science sans conscience n'est que ruine de l'âme». Así escribía Rabelais en el Pantagruel (1532), anticipando de forma literaria, en el discurso que Gargantúa dirige a Pantagruel, un dilema aún vigente: el del poder del conocimiento sin responsabilidad ética.

[3] Que la evolución técnica reclame normas, no como obstáculos para su expansión, sino como condición de su legitimidad, es una idea que se ha beneficiado ampliamente de la reflexión de N. IRTI-E. SEVERINO, *Dialogo su diritto e tecnica*, Roma-Bari, 2001; N. IRTI, *Nichilismo giuridico*, Roma-Bari, 2004; ID., *Il diritto nell'età della tecnica*, Napoli, 2007; y ID., *Diritto e tecnica (in dialogo con Emanuele Severino e Luigi Mengoni)*, en *Lo stato*, 2016, p. 35 ss.

[4] Hasta el punto de concebir una relación de "subyugación" conscientemente instrumental: cf. A.M. BENEDETTI, *Contratto, algoritmi e diritto civile transnazionale: cinque questioni e due scenari*, en *Riv. dir. civ.*, 2021, p. 411 ss.; M. ZACCHEO, *Panopticon*, en M. PROTO (a cura di),

Como profesor de Derecho civil, pero aún antes como estudioso y como hombre de ciencia, me comprometo a no defender una posición de principio, sino a ofrecer una lectura crítica y al mismo tiempo propositiva.

Porque estoy convencido – y lo diré sin vacilar – de que la inteligencia artificial, lejos de representar una amenaza para la producción científica, puede constituir una oportunidad extraordinaria.

Sin embargo, bajo una condición: que permanezcamos fieles a los principios que hacen que la ciencia sea digna de ese nombre. Originalidad. Rigor metodológico. Innovación. Y, sobre todo, responsabilidad humana [5].

2. Ciencia, responsabilidad y técnica: el principio de autorresponsabilidad

En mi labor académica – y ahora también como director de una revista internacional (*AI Law - International Review of Artificial Intelligence Law*) [6] – siempre he considerado la ciencia como un ejercicio de libertad no arbitraria.

Una libertad intelectual que encuentra en la metodología, en la transparencia y en la verificabilidad no obstáculos, sino instrumentos de regeneración continua.

La ciencia, en suma, nunca es pura espontaneidad, ni mera repetición. Es, cada vez, un acto deliberado de construcción del conocimiento, sostenido por criterios compartidos.

En una época en la que la técnica penetra en los propios mecanismos de la

Intelligenza artificiale e rapporti bancari. Umano e non umano nelle relazioni tra intermediari e clienti, Pisa, 2024, p. 16. Es más, para E. MINERVINI, *Note brevi sull'intelligenza artificiale*, en *AI Law*, 2025, p. 190, «l'intelligenza artificiale, quale fenomeno rivoluzionario, rappresenta forse il laboratorio più avanzato del diritto civile e degli studi civilistici».

[5] Se requiere un principio ético concebido para una época en la que la técnica posee un poder irreversible y global, y en esta clave de lectura, centrada en el vínculo entre la ética del conocimiento y la responsabilidad, hago mío y asumo el pensamiento de H. JONAS, *Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation*, Frankfurt am Main, 1979, p. 36: «Handle so, daß die Wirkungen deiner Handlung verträglich sind mit der Permanenz echten menschlichen Lebens auf Erden». Sobre este punto, véanse los estudios de J. SCHUBERT, *Das «Prinzip Verantwortung» als verfassungsstaatliches Rechtsprinzip. Rechtsphilosophische und verfassungsrechtliche Betrachtungen zur Verantwortungsethik von Hans Jonas*, Baden-Baden, 1998, quien describe dicha responsabilidad como *Schlüssel zu einer neu-orientierten Staatstheorie, die sich nicht nur mit der Verankerung der Verantwortungsethik im Verfassungsrecht beschäftigt, sondern gleichfalls eine Theorie der politischen Verantwortung enthält und ebenso den Begriff der Bürgerverantwortung im Verfassungsstaat beeinfluß*.

[6] https://www.reviewofailaw.com.

investigación, resulta esencial reafirmar un principio fundante: aquel que yo denomino "autorresponsabilidad epistémica" [7].

Quien elabora un artículo científico, quien firma una tesis doctoral, quien redacta un ensayo o una monografía, asume la responsabilidad no solo de lo que escribe, sino también de los instrumentos que ha decidido emplear para escribirlo [8].

Sobre este punto fundé el código ético de la revista *AI Law* [9].

No se trata – lo repito – de prohibir la inteligencia artificial, sino de legitimarla únicamente a condición de que se declare su utilizo, se controle, se integre y verifique. En una palabra: se asuma humanamente.

Este es el mismo espíritu que anima las Directrices elaboradas por la Comisión Europea en 2024 para el uso responsable de la IA generativa en la investigación. En ellas se afirma con claridad: la inteligencia artificial se puede utilizar, pero siempre bajo control humano, de manera transparente y responsable, y nunca como fuente autónoma de verdad o como sujeto epistémico [10].

La centralidad del sujeto humano – y de su juicio – no constituye un residuo antropocéntrico. Es, más bien, la condición de posibilidad de todo conocimiento auténtico.

---

[7] Por su innegable derivación conceptual, pero también para evaluar su fidelidad al pensamiento del autor, véase L. CODE, *Epistemic Responsibility*, Hanover, 1987, quien describe su vínculo con la ética del conocimiento y con la autenticidad del acto cognitivo.

[8] Esta orientación, que atribuye al sujeto humano un papel insustituible de control, discernimiento y responsabilidad, es perfectamente coherente con lo afirmado en el ámbito jurídico por C. PERLINGIERI, *Intelligenza Artificiale tra princípi e regole*, en *AI Law*, 2025, p. 42, para quien «il fondamento della regolamentazione è insito innanzitutto nel principio del personalismo che si traduce nell'adozione di un approccio dell'IA al servizio dell'uomo [...] La logica antropocentrica acquista effettività, infatti, soltanto se integrata da un risk approach il quale [...] mira ad assicurare l'affidabilità dei sistemi». Sobre el papel fundacional del principio personalista insiste con fuerza también V. VERDICCHIO, *IA e diritto: una introduzione*, *ivi*, p. 152 ss.; pero para una crítica decidida, en concreto, de este *risk approach*, cf. M. ZACCHEO, *ob. cit*., p. 24: «Il criterio del rischio, che peraltro inopportunamente non è definito, potendosi del termine predicare una pluralità di significati, necessariamente va considerato nella sua accezione più classica, come il pericolo di un male. Nonostante l'ottimismo legislativo, il male assoluto, e dunque vietato, è rintracciato nelle cd. Pratiche vietate, indicate nell'art. 5 del Regolamento, definite rischio grave; il male è, invece, individuato nei sistemi cd. ad Alto rischio dove l'obiettivo è fissare regole specifiche per sistemi di I.A. «che creano un rischio alto per la salute e la sicurezza o per i diritti fondamentali delle persone fisiche»».

[9] https://www.reviewofailaw.com.

[10] La IA no es técnicamente un intérprete, sino un objeto que elabora determinadas funciones mediante algoritmos, careciendo de la capacidad de comprensión, que es, en cambio, una propiedad no algorítmica fundamental de la conciencia. Cf. M. ZACCHEO, *ob. cit*., p. 13 ss.

3. Uso responsable de la IA: una gramática común

El éxito de la inteligencia artificial no dependerá únicamente de su potencia técnica. Dependerá, sobre todo, de la madurez del contexto que la acoge [11].

Esto vale en todos los ámbitos, pero de modo muy particular en la producción científica, donde lo que está en juego no solo es la eficiencia, sino la calidad del conocimiento y la confianza de la comunidad que lo reconoce [12].

En los dos últimos años hemos asistido a un verdadero florecimiento de guías de buenas prácticas, políticas académicas y códigos éticos, orientados a dar forma, contenido y medida a esta nueva madurez.

No se trata de vínculos impuestos desde arriba, sino – cuando están bien concebidos – de gramáticas compartidas, que hacen posible un uso virtuoso de la inteligencia artificial: un uso que faculta sin absolver, que facilita sin eximir de responsabilidad.

Entre los documentos más autorizados, deseo recordar – sin pretensión de exhaustividad, pero con plena convicción – las *Living Guidelines* [13], que invocan los principios de fiabilidad, honestidad, respeto y responsabilidad; las *Ethics Guidelines for Trustworthy AI* [14], que fijan requisitos operativos fundamentales como la transparencia, la supervisión humana, la equidad y la explicabilidad; y las directrices italianas elaboradas por las Universidades de Ca' Foscari, Bolonia y Siena, que supieron traducir dichos principios en instrucciones concretas y practicables.

Todos estos textos convergen, en el fondo, en algunos puntos esenciales.

Podemos considerarlos la gramática mínima del uso responsable de la inteligencia artificial en la investigación, que paso ahora a enunciar.

a) *Legitimidad condicionada*

La inteligencia artificial queda legítima únicamente como instrumento de apoyo. Nunca como fuente autónoma de contenido científico. Su legitimidad

[11] La interacción entre ciencia y tecnología, por lo demás, no es unidireccional, sino una dinámica recíproca, en la que incluso el Derecho se ve transformado por los instrumentos que intenta regular. Cf. N. MACCABIANI, *Data Processing and Algorithms: A Constitutionally Oriented AI?*, en *AI Law*, 2025, p. 7: «Law not only strives to regulate new technologies but, in turn, is shaped by technological advancements in a twofold way: for its content and for the deployed regulatory techniques that evolve according to the different types of technologies and implied risks».

[12] Cf. S. MIRABELLO, *«Artificial Intelligence, Development and Evaluation of Scientific Knowledge». Report on a Seminar*, en *AI Law*, 2025, p. 196 ss.

[13] Comisión Europea (2024): https://european-research-area.ec.europa.eu.

[14] Comisión Europea (2019): https://digital-strategy.ec.europa.eu.

depende de que se emplee de forma consciente, crítica y subordinada, como un auxilio técnico, no como un sustituto del pensamiento.

b) *Transparencia en el uso*

Todo uso significativo de la IA – en la selección de las fuentes, en la estructuración de la argumentación, en la generación lingüística – se ha de declarar explícitamente [15]. Como se afirma también en el código ético de *AI Law*, esto vale igualmente para operaciones menos visibles: la elaboración de la bibliografía, el análisis preliminar de los datos o la redacción del *abstract*. El lector – la comunidad – tiene derecho a saber qué parte es fruto del autor y cuál está mediada por una herramienta automatizada.

c) *Control humano*

El resultado generado por la IA se ha de verificar punto por punto, especialmente en lo que respecta a: la exactitud de las citas; la coherencia lógica; la detección de posibles sesgos; el respeto de los derechos de autor y de los datos sensibles. El control humano no es opcional: es fundacional [16]. En él se concreta el principio de exactitud, reconocido por las políticas universitarias actualmente vigentes en Italia en relación con la producción científica. El análisis humano sirve para prevenir errores o imprecisiones, asegurando que la información esencial sea correctamente presentada, sin distorsionar su auténtico significado [17].

La verificación de la correcta interpretación de las fuentes constituye un paso ineludible para garantizar la continuidad de una investigación "sana": síntesis del pensamiento pasado y punto de partida para una revisión innovadora en el presente.

[15] ALLEA, *European Code of Conduct for Research Integrity*, 2023, § 2.3.1: «Researchers must disclose the use of generative tools and specify how they were employed in the research process».

[16] El algoritmo muestra, pero no explica, y su fuerza consiste precisamente en hacerse aceptar como neutro, objetivo e impersonal, cuando en realidad sustituye la decisión por el cálculo, el juicio por el procedimiento y la conciencia por la instrucción. Este es el sentido general que extraigo de M. ZACCHEO, *ob. cit*., p. 13 ss.

[17] Cf. EUROPEAN COMMISSION, *Living Guidelines on the Responsible Use of Generative AI in Research*, Brussels, 2024, p. 6: «Human oversight is a structural element in ensuring the accuracy, integrity and trustworthiness of research practices assisted by AI».

d) *Exclusión de la autoría de la IA*

Ningún sistema de inteligencia artificial puede ser indicado como autor ni coautor de un trabajo científico. La ciencia reconoce como autor a un sujeto capaz de intención, responsabilidad y juicio crítico. Y la inteligencia artificial, por definición, carece de todo ello.

e) *Limitación en la* peer review

La inteligencia artificial puede ser utilizada por los revisores solo con fines formales: para controles léxicos, sintácticos o bibliográficos. Pero nunca puede formular un juicio sustantivo sobre el valor científico de un texto [18]. Porque juzgar implica comprender el mérito. Y el mérito exige experiencia, contexto, confrontación y cultura.

Estos criterios no pueden permanecer implícitos. Deben ser formalizados en los códigos éticos, enseñados en las escuelas de doctorado, exigidos por las revistas científicas y evaluados en los procedimientos académicos.

Solo así evitaremos los dos extremos: por un lado, el uso despreocupado y exento de responsabilidad de la IA; por el otro, su exclusión prejuiciosa, fruto del temor o de la incomprensión.

La inteligencia artificial no es un problema que hay que resolver. Es, más bien, una responsabilidad que tenemos que asumir.

4. Hacia una nueva metodología de la investigación

Si hay un punto sobre el cual deseo ser absolutamente claro, es este: la inteligencia artificial no constituye una amenaza para la ciencia, sino un desafío a su evolución metodológica.

Un desafío que puede – si se asume con inteligencia, rigor y sentido crítico – convertirse en una extraordinaria oportunidad de regeneración.

El acto de investigar – desde la intuición inicial hasta la comunicación de los

[18] «Gli algoritmi che elaborano i dati e prendono "decisioni" sono spesso di una complessità tale da risultare incomprensibili non solo per i comuni cittadini, ma anche per molti esperti. Questa "scatola nera" algoritmica solleva serie questioni di trasparenza e responsabilità: come possiamo contestare o anche solo comprendere decisioni che influenzano profondamente le nostre vite se non abbiamo accesso ai meccanismi che le generano?». Con esta interrogante, S. PIETROPAOLI, *Dalla sorveglianza al controllo: la parabola della governamentalità algoritmica*, en *Riv. it. inf. dir.*, 2-2024, p. 29, plantea cuestiones que no se refieren directamente solo a la evaluación de la investigación, sino que la implican, a mi juicio, en un contexto sistémico general particularmente adecuado y mucho más amplio.

resultados – siempre ha estado atravesado, transformado y potenciado por tecnologías cognitivas [19]: desde la imprenta hasta el ordenador, desde el índice bibliográfico hasta el motor de búsqueda, desde el procesador de textos hasta las bases de datos jurídicas.

La inteligencia artificial generativa es la última – y quizá la más invasiva – de estas transformaciones.

No es una novedad absoluta, sino un salto de escala, de acceso y de velocidad.

Y, al verlo con lucidez, hay que reconocer que la IA – cuando se utiliza con criterio – ya hoy puede mejorar la calidad de la producción científica.

Mejorarla no tanto en los contenidos, sino en los procesos: en las operaciones de búsqueda, de organización, de formulación. A continuación, deseo ilustrar algunos de los ámbitos en los que esta contribución resulta más evidente:

a) *Selección y gestión de las fuentes*

Gracias a herramientas como Elicit, Research Rabbit o Scite, el investigador puede identificar con rapidez las fuentes más pertinentes sobre un determinado tema; explorar las conexiones entre artículos, autores y corrientes doctrinales; y obtener visualizaciones de citas o redes temáticas.

Nada de esto sustituye el juicio crítico, sino que lo potencia: hace que la investigación sea más informada, más amplia y más estratégica.

b) *Traducción y accesibilidad lingüística*

Con traductores avanzados como DeepL o Google Translate en sus versiones mejoradas, es hoy posible traducir textos académicos complejos con un grado notable de fidelidad; redactar resúmenes o presentaciones en otros idiomas; comparar versiones lingüísticas de una misma fuente normativa.

Ello favorece la internacionalización de la investigación, especialmente para quienes trabajan en contextos multilingües o interdisciplinarios. El acceso a las fuentes del conocimiento deja así de encontrarse condicionado por el idioma.

[19] «We are increasingly delegating epistemic tasks to artificial agents, and in so doing, we are undergoing a fourth revolution in our self-understanding»: L. FLORIDI, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford, 2014, p. 45, para quien, esencialmente, la cuarta revolución cognitiva transfiere poder epistémico de la mente humana al artefacto informacional.

c) *Redacción y revisión de textos*

Aplicaciones como Grammarly, Writefull, Trinka o ChatGPT pueden contribuir a mejorar la fluidez y la corrección estilística; sugerir alternativas lingüísticas más precisas; uniformar citas, párrafos, notas y títulos.

En este ámbito, la IA no es un autor, sino un asistente editorial. Ahorra tiempo, evita errores y logra mejorar la legibilidad del producto científico.

d) *Verificación de la coherencia lógica*

Algunas herramientas – aún en fase de desarrollo – permiten ya verificar la coherencia argumentativa de un texto, señalar pasajes redundantes o contradictorios y sugerir una reorganización más lineal de las tesis.

En el fondo, se trata de simular la voz de un primer lector crítico: esa voz que todo autor desea escuchar, pero que con frecuencia falta en las etapas iniciales del trabajo.

e) *Construcción asistida de aparatos informativos*

Es cada vez más común el uso de la IA para generar automáticamente bibliografías, repertorios normativos o jurisprudenciales; clasificar fuentes y notas; o actualizar en tiempo real los referentes legales y doctrinales.

La IA se convierte así en un curador técnico, garante de orden, exhaustividad y actualización.

Pero conviene advertir: nada de esto garantiza, por sí solo, la calidad del contenido.

El riesgo – real y actual – es confundir la forma con el fondo, la fluidez con la originalidad, la eficiencia con la autenticidad.

Hoy es posible, mucho más que en el pasado, publicar un artículo formalmente perfecto que, sin embargo, no dice nada; un texto impecable en su estructura, pero vacío en su pensamiento [20].

Por ello creo que la inteligencia artificial, si se utiliza con conciencia y medida, puede liberar tiempo y energía, permitiendo a los investigadores concentrarse en lo que realmente importa: las ideas, las tesis, las objeciones, las innovaciones.

Es una cuestión de papel que se la atribuye a la IA: hay que entenderla como asistente, no como autor. Pero también es una cuestión de disciplina: no para castigar el uso de la tecnología, sino para exigir la responsabilidad científica, que sigue siendo – y seguirá siendo siempre – humana.

---

[20] La idea de que la elegancia formal puede enmascarar el vacío del pensamiento está en el centro de la reflexión de J. HABERMAS, *Erkenntnis und Interesse*, Frankfurt am Main, 1968, quien advierte contra el tecnicismo desvinculado de la función crítica del conocimiento.

5. El papel del Derecho y de las instituciones académicas

El uso de la inteligencia artificial en la producción científica no puede quedar librado al azar, ni mucho menos a la libre iniciativa individual.

No se trata – lo repito – de un mero instrumento técnico, sino de un elemento que incide directamente en la epistemología, la responsabilidad y la legitimación del conocimiento [21].

Por ello, hoy el Derecho, las universidades, las agencias de evaluación y las revistas científicas no pueden limitarse a observar. Deben asumir un papel activo, previsor y crítico. Deben convertirse – en sentido pleno – en instituciones reguladoras de la transformación.

a) *El Derecho como arte de la medida*

Como jurista, estoy habituado a concebir la norma no como un obstáculo, sino como la condición de posibilidad de una libertad responsable.

En el caso de la inteligencia artificial, lo que falta no es la tecnología, sino una medida compartida de su uso legítimo.

Con este espíritu, la Unión Europea ha acompañado al conocido AI Act con una serie de instrumentos normativos y meta-normativos, entre los cuales pueden destacarse las *Living Guidelines on the Responsible Use of Generative AI in Research*.

Estas directrices no prohíben, no reprimen, no bloquean. Pero reconducen el uso de la IA a un marco ético y operativo bien definido, fundado en cuatro obligaciones fundamentales: la transparencia en su utilización; la verificación de los resultados; la formación continua del investigador; y la declaración explícita en las publicaciones y proyectos.

[21] Como ha observado acertadamente Maria Rosaria Ferrarese, la llamada "normatividad tecnológica" se presenta con características de rigidez y opacidad: «non solo non può esplicitare la *ratio* dei suoi comandi, perché questi sono contenuti negli algoritmi, e dunque in una lingua non enunciabile, ma i suoi comandi ricevono esecuzione in tempo reale, in maniera automatica e senza bisogno di esecutori»: cf. M.R. FERRARESE, *Normatività in competizione: economia e tecnologie digitali erodono lo spazio del diritto?*, en G. AZZARITI, S. BAGNI, M. CARDUCCI e A. SOMMA (a cura di), *Mercato, istituzioni e regole. Sul difficile dialogo tra diritto ed economia*, Roma, 2025, p. 110, quien concluye: «Come tale, la normatività tecnologica non è mai del tutto innocente, come tende in genere a presentarsi, in quanto è condizionata proprio da quelle regole invisibili che sono interne agli algoritmi, e che possono incidere anche sulla gamma dei diritti, sia nella sfera privata che in quella pubblica. Tutto il tema dell'intelligenza artificiale e delle sue prestazioni ed evoluzioni rimanda al problema di questa invisibile e potente capacità performativa di cui essa è impregnata».

Pero sobre la «natura politica della tecnologia», considero decisiva la observación de L. CORSO, *Breve tassonomia dei rapporti fra Intelligenza Artificiale, etica e diritto. Tre questioni*, en *AI Law*, 2025, p. 222.

En otras palabras, la IA queda disciplinada como cualquier otra tecnología jurídicamente relevante: no como sujeto, sino como objeto de responsabilidad y de control.

b) *Las universidades como laboratorios normativos*

Junto a las normas europeas se multiplican – y con razón – las políticas universitarias.

En Italia, en particular, las directrices de Ca' Foscari, Bolonia y Siena ofrecen modelos complementarios e inteligentes.

Siena adopta un enfoque formativo y pedagógico, basado en la confianza y en la conciencia [22].

Ca' Foscari construye una tipología analítica de los casos de uso, distinguiendo claramente lo que es lícito, lo que es útil y lo que resulta riesgoso [23].

Bolonia, por su parte, promueve un comité académico para la IA que potencia el diálogo entre la comunidad científica, las instituciones y la sociedad. Proporciona orientaciones concretas para trasladar las directrices teóricas a contextos operativos específicos y, al mismo tiempo, estimula la reflexión sobre posibles aplicaciones futuras. Incentiva un nuevo enfoque para mejorar la actividad investigadora, siempre que ello ocurra en el respeto de la calidad, de la ética y de la dignidad de la persona [24].

Los tres modelos, aunque diferentes, convergen en un punto esencial: la universidad ya no puede ser un sujeto pasivo. Debe convertirse en laboratorio de normatividad, capaz de acompañar, interpretar y anticipar.

c) *Las revistas científicas como garantes de fiabilidad y custodias del "bienestar" de la investigación*

Un papel decisivo corresponde —y seguirá correspondiendo— a las revistas científicas, lugar donde el conocimiento se produce, se evalúa y se reconoce.

Las revistas son – o deberían ser – baluartes de fiabilidad epistemológica, y por lo tanto también de responsabilidad [25].

---

[22] L'ateneo stabilisce una policy che definisce l'utilizzo di ChatGPT e altri Large Language Models (LLM): https://www.unisi.it.

[23] Linee Guida per l'uso responsabile dell'IA generativa nel campo della didattica e dell'apprendimento: https://www.unive.it; Linee guida per l'uso responsabile dell'intelligenza artificiale a fini di ricerca: https://www.unive.it.

[24] Policy di Ateneo per un uso etico e responsabile dell'Intelligenza Artificiale Generativa nelle attività di didattica e ricerca: https://www.unibo.it.

[25] Cfr. COPE (Committee on Publication Ethics), Code of Conduct and Best Practice Guide-

Con esta convicción fundé la revista *AI Law - International Review of Artificial Intelligence Law*, con el propósito de crear un espacio no solo de reflexión jurídica, sino también de práctica académica regulada, capaz de acoger la tecnología sin renunciar al rigor metodológico.

El Código Ético que hemos adoptado – inspirado en las directrices del COPE [26], pero adaptado a las especificidades de la IA – establece con precisión que la inteligencia artificial solo se puede utilizar como apoyo, nunca como sustituto de la subjetividad científica; que todo uso significativo hay que declarlo explícitamente, con indicación del instrumento, la versión y los *prompts* empleados; que la contribución humana tendrá que ser siempre reconocible, verificable y metodológicamente fundada; y que la *peer review* puede recurrir a herramientas de IA únicamente para aspectos formales (coherencia, bibliografía), nunca para evaluaciones de fondo.

En este enfoque subyace una convicción profunda: regular el uso de la inteligencia artificial no significa obstaculizarla, sino darle una forma jurídica y científica, para que pueda integrarse sin deformar aquello que la ciencia debe seguir siendo: una responsabilidad racional y verificable ante la "verdad" [27].

A las revistas científicas, en particular, les corresponde la tarea de salvaguardar escrupulosamente la "salud" del conocimiento, al cual el uso de tecnologías avanzadas puede aportar beneficios, siempre que esté orientado al cuidado de los componentes esenciales del núcleo inalterable de la investigación científica.

Frente a la inteligencia artificial, el investigador debe ser "escrupulosamente crítico" y "críticamente constructivo" [28].

El uso de herramientas tecnológicas en este sentido revaloriza la aportación

lines for Journal Editors, 2011, quien define las revistas como «custodi della qualità e dell'integrità del sapere scientifico».

[26] https://publicationethics.org.

[27] No puedo ocultar la incomodidad que siempre siento ante esta palabra, y también por esta razón quisiera aclarar que aquí la entiendo en el sentido más específico y limitado que extraigo de la lectura de M. FOUCAULT, *L'ordre du discours*, Paris, 1971, p. 32: «Le discours scientifique est bien un discours ordonné par une certaine règle de construction, par un certain jeu de contraintes; il n'est pas simplement celui qui dit ce qui est vrai, mais celui qui se conforme à des règles de vérité».

[28] Cf. L. CORSO, *ob. cit*., p. 222: «fino a che punto e in che misura, l'IA, che di certo ci sgrava da moltissimi compiti demandati alla macchina, mantiene intatta la capacità riflessiva dell'uomo, che è poi alla base della sua capacità morale? Questa capacità è l'unico strumento che può contrastare il nostro soccombere di fronte alla macchina: a chi denuncia la natura politica della tecnologia, si può obiettare che nella misura in cui venga preservata la possibilità del pensiero dell'uomo, anche sugli usi che vengono fatti dalla macchina, il rischio di un collasso morale è scongiurato».

personal del investigador, que adquiere formas peculiares y originales según el tipo de interacción escogida, los criterios aplicados, la formulación del *prompt*, la selección del conjunto de datos de entrenamiento, las instrucciones de refinamiento y muchos otros factores que – adecuadamente o incluso fortuitamente configurados – pueden conducir a resultados innovadores e inesperados. Todo ello, sin perder de vista el riesgo moral inherente a la posible pérdida de control pleno sobre el rigor metodológico o incluso sobre la conciencia epistemológica [29].

No obstante, estoy convencido de que ningún investigador renunciaría, por principio, a un resultado auténticamente original o innovador solo porque la caja negra de la IA le impida conocer con certeza el modo en que dicho resultado fue alcanzado.

6. Una voz personal: ética de la prudencia, ética de la confianza, ética del cuidado

Quisiera ahora cerrar – o mejor dicho, empezar a cerrar – esta intervención con una reflexión más personal.

No hablo aquí como director de una revista, ni como profesor de Derecho civil, sino como hombre de ciencia, que ha vivido cuarenta años de universidad no como una profesión, sino como una vocación.

He visto cambiar muchas cosas en la manera de investigar, de escribir, de enseñar, de evaluar. He visto transformarse los instrumentos, los lenguajes y las formas de visibilidad.

Y, sin embargo, en todo ello, hay algo que no ha cambiado – y que no debe cambiar: la ética de la investigación, el compromiso serio – no retórico, sino radical – de garantizar que lo que se dice en nombre de la ciencia sea fruto de un pensamiento auténtico, de un método riguroso y de una responsabilidad personal.

Hoy, con la inteligencia artificial, nos hallamos ante una prueba de madurez colectiva.

Se nos pide aceptar instrumentos poderosos – a veces seductores – sin dejar de pensar; acelerar sin perder la dirección; mejorar sin deshumanizar [30].

---

[29] En esta dimensión, a mi juicio, las palabras de H. JONAS, *ob. cit*., p. 7, alcanzan hoy el núcleo mismo del problema: «Daß wir imstande sind zu tun, was wir nicht imstande sind vorauszusehen, und daß wir imstande sind vorauszusehen, was wir nicht imstande sind zu tun, sind die neuen Tatsachen, mit denen die Ethik sich abfinden muß».

[30] De este modo, a mi juicio, debe acogerse y valorarse el anhelo expresado por V. VERDICCHIO, *ob. cit*., p. 155, «ad approfondire la conoscenza della *materia tecnologica* che il diritto ambisce a regolare».

Creo que esto es posible. Pero bajo tres condiciones.

La primera es la *ética de la prudencia*. No entendida como desconfianza, sino como capacidad de interrogarse. De no ceder al automatismo de lo nuevo solo por ser nuevo. De recordar que escribir un texto siempre corresponde a un acto de responsabilidad.

La segunda es la *ética de la confianza*. Confianza en que las tecnologías – si se utilizan correctamente – pueden ampliar nuestra capacidad de investigación. Confianza en que regular no significa reprimir, sino hacer posible de manera justa.

No necesitamos una academia encerrada en el temor al futuro, ni una academia eufórica y sin medida.

Necesitamos – eso sí – una comunidad capaz de volver a las preguntas fundamentales: ¿qué significa firmar un texto?, ¿qué significa sostener una tesis?, ¿qué significa ser responsable de una idea?

Mi respuesta es sencilla, quizá también antigua: no importa la velocidad con la que escribamos, sino cuánto de nosotros está presente en lo que escribimos.

No importa si usamos una, diez o cien herramientas, sino si lo que decimos procede de una voz capaz de responder por su propio decir [31].

Fue también por esta razón que quise fundar *AI Law*: no para vigilar, sino para orientar; no para reprimir, sino para dar forma ética y jurídica a la apertura; no para decir "no" a la inteligencia artificial, sino para afirmar un "sí" consciente, maduro y regulado a una tecnología que puede potenciar la ciencia sin alterar su vocación veritativa.

La tercera es la *ética del cuidado*. Como miembros de una comunidad académica, tenemos el deber de preservar la dignidad de nuestro trabajo, que es y sigue siendo, ante todo, investigación científica: no delegable ni imputable a entidades incapaces de un enfoque crítico autónomo.

Todos estamos obligados a garantizar el respeto del "derecho a la salud" de la investigación. Solo los valores académicos de integridad y reconocimiento pueden asegurar una convivencia entre estudiosos basada en la honestidad.

Si el fin último de nuestra comunidad es el avance de la ciencia, habrá que respetar el trabajo de quienes nos precedieron y saber rendir cuentas de él, incluso si es para refutarlo y superarlo. Del pasado, en efecto, a veces es bueno liberarse, pero conscientemente, no por ignorancia [32].

---

[31] En el sentido explicitado por el razonamiento de H. JONAS, *ob. cit*., p. 228, che così sintetizza il dovere di assumere responsabilità epistemica: «In der Verantwortung begegnet uns das Phänomen der Pflicht in seiner ursprünglichsten Form».

[32] Quisiera que esta afirmación permaneciera inspirada en la reflexión epistemológica de P. RICŒUR, *La mémoire, l'histoire, l'oubli*, Paris, 2000, p. 579: «En effet, avant l'abus, il y a

La producción científica más reciente, víctima de mecanismos perversos [33], crece en cantidad, pero muestra un preocupante acortamiento de su horizonte temporal de conocimiento. Encuentra en la inteligencia artificial un aliado natural, pero también un abrazo mortal, pues la IA está sujeta al límite de la digitalización del saber [34], que le permite entrenarse solo sobre bases de datos. De ahí el melancólico contraste con aquellas bibliotecas de papel que van cubriéndose de polvo, porque muchos jóvenes nunca las han visitado [35].

La ética del cuidado plantea, por tanto, también un problema generacional. No hay progreso en el uso malsano o irreflexivo de una herramienta tecnológica ontológicamente limitada al presente y a su inmediatez [36]. Por eso merece ser saludado con entusiasmo el experimento llevado a cabo en la Universidad de

---

l'usage, à savoir le caractère inéluctablement sélectif du récit. Si on ne peut se souvenir de tout, on ne peut pas non plus tout raconter. L'idée de récit exhaustif est une idée performativement impossible. Le récit comporte par nécessité une dimension sélective. Nous touchons ici au rapport étroit entre mémoire déclarative, narrativité, témoignage, représentation figurée du passé historique. Comme nous le notions alors, l'idéologisation de la mémoire est rendue possible par les ressources de variation qu'offre le travail de configuration narrative. Les stratégies de l'oubli se greffent directement sur ce travail de configuration».

Que una teoría solo pueda ser rechazada después de haber sido puesta a prueba con rigor es lo que explica, con la agudeza que lo ha hecho célebre, K. POPPER, *The Logic of Scientific Discovery*, London, 1934. Tampoco la refutación, al fin y al cabo, es más que un acto cognitivo, no un juicio arbitrario.

Prefiero confiar la conclusión de esta nota a la idea de que, para superar una doctrina, primero es necesario comprenderla; y para comprenderla, es preciso haberla estudiado con respeto. Desde esta perspectiva, me viene a la mente, ante todo, el peso moral y científico, así como el coraje de N. BOBBIO, *Politica e cultura*, Torino, 1955; ID., *Profilo ideologico del Novecento*, Milano, 1990.

[33] Comparto plenamente (y recomiendo encarecidamente la lectura de) F. D'ALESSANDRO, *Rumore (ovvero come la pseudo-scienza stia soffocando la scienza vera)*, en *Riv. dir. civ.*, 2023, p. 96 ss., de quien debería valorarse especialmente la sugerencia contenida en la letra d) de la p. 115.

[34] El conocimiento mediado por algoritmos corre el riesgo de encerrar el intelecto en un ecosistema informativo autorreferencial. Cfr. E. PARISER, *The Filter Bubble. What the Internet Is Hiding from You*, New York, 2011, p. 86: «If it hasn't been digitized, it doesn't exist in the filter bubble».

[35] Con una pequeña licencia, coloco aquí, en esta posición, la referencia a J.-C. CARRIÈRE-U. ECO, *Non sperate di liberarvi dei libri*, a cura di J.-Ph. de Tonnac, trad. de A.M. Lorusso, Milano, 2009.

[36] Si se permitiera una libre adaptación de una célebre máxima apoftegmática, casi podría formularse un *axioma technocraticum*: «quod non est in Internet aut in Big Data, non est in mundo». Volviendo a un tono más serio, sería, en cualquier caso, sumamente aconsejable, también y sobre todo en este contexto, mantener siempre presente aquella célebre enseñanza socrática sobre el saber.

Milán, donde se ha desarrollado AILexA, un algoritmo concebido para trabajar sobre el patrimonio sapiencial de la *lex Aquilia* [37].

El esfuerzo de un joven investigador infunde esperanza e indica el camino a seguir.

Hasta hoy, sin embargo, las antiguas colecciones de jurisprudencia no digitalizadas deben seguir siendo estudiadas, precisamente – si se prefiere – porque la inteligencia artificial, por ahora, no puede servirse de ellas.

7. La IA en la ciencia: una ocasión de regeneración

Podríamos preguntarnos – y concluir así – si la inteligencia artificial cambiará realmente todo. La respuesta, a mi juicio, no es unívoca.

La IA cambiará mucho, quizá casi todo. Pero solo si nosotros lo permitimos. Transformará los tiempos de la investigación, las formas de escritura, los instrumentos de evaluación, las modalidades de la comunicación científica. Pero nunca podrá sustituir aquello que hace que un texto sea verdaderamente "científico": la invención de la pregunta, la claridad de la intuición, el coraje de una tesis original, el deber de responder por el propio método [38].

Jamás podrá generar lo que aún no ha sido dicho, porque la ciencia no solo es predicción: es riesgo, es hipótesis, es exposición [39].

---

[37] Véase R. PERANI, *AILexA. L'intelligenza artificiale applicata alla lex Aquilia*, Milano, 2024; F. ADDIS, *Einleitung in das Kolloquium an der Universität Bern (20.-21. März 2024): «Automatisierung von juristischen Entscheidungen. Die künstliche Intelligenz am Beispiel der römischrechtlichen Kasuistik»*, en I. Fargnoli (Hrsg.), *Automatisierung von juristischen Entscheidungen. Die künstliche Intelligenz am Beispiel der römischrechtlichen Kasuistik*, Milano, 2024, p. 17 ss.; R. PERANI, *Revisiting the Digest through scripts and algorithms*, *ivi*, p. 81 ss.; y antes aún ID., *Intelligenza artificiale e* Digesta Iustiniani. *La casistica romana per un orientamento nella risoluzione automatizzata delle controversie*, Milano, 2023; obra sobre la cual, cf. S. SCHIPANI, *Presentazione del volume di Renato Perani,* Intelligenza artificiale e *Digesta Iustiniani*. La casistica romana per un orientamento nella risoluzione automatizzata delle controversie *(Milano, 2023) e invito alla attenzione al* cottidie ius in melius produci, en *AI Law*, 2025, p. 157 ss.; y véanse también las precisiones de R. PERANI, *AILexA e la sfida della risoluzione automatizzata delle controversie*, *ivi*, p. 183 ss.

[38] Es preciso contrarrestar esa especie de reduccionismo intelectual que resulta, especialmente en el ámbito de la investigación, del uso de tecnologías automatizadas. Es necesario, en particular, evitar esa decadencia cognitiva que B. STIEGLER, *La società automatica*, 1, *L'avvenire del lavoro*, trad. y ed. por S. Baranzoni, I. Pelgreffi y P. Vignola, Milano, 2019, prefigura con tonos sombríos y describe de manera realmente preocupante en su amplia producción filosófica.

[39] Cf. A. MASALA, *Algorithms and the Abuse of Reason. Reflections from Friedrich A. von Hayek's Theory of Knowledge*, en *Metábasis.it*, 38/2024, p. 194, donde se pone de relieve que

Y aquí – en este espacio entre técnica y creación – nos sale al encuentro un pensamiento antiguo y poderoso.

Hubo un momento originario de la cultura occidental en el que la forma y el caos, la medida y el éxtasis, lo apolíneo y lo dionisíaco se unieron en una tensión fecunda, de la cual nació la tragedia griega.

Dos impulsos profundos: uno hacia la claridad, el orden, la simetría; el otro hacia la intensidad, la ruptura, el vértigo. Separados, se atrofian. Unidos, generan arte. Generan verdad. Generan ciencia.

Hoy, frente a la inteligencia artificial, tal vez nos encontremos de nuevo dentro de esa misma tensión. Por un lado, el impulso apolíneo del lenguaje generado, de la coherencia formal, de la claridad predictiva. Por otro, el impulso dionisíaco de la investigación viva, de la intuición radical, de la duda, del error, del riesgo cognitivo.

Los necesitamos ambos. Si la IA nos ofrece lo apolíneo, nos corresponde a nosotros preservar lo dionisíaco. No para oponernos a la técnica, sino para salvar su sentido. Porque un texto puede ser formalmente perfecto y, sin embargo, carecer de vida. Una tesis puede ser lógicamente coherente y, sin embargo, no interrogar nada.

El peligro de la inteligencia artificial no es la falsedad, sino la verdad estéril: la verdad sin lucha, sin voz, sin historia.

La tragedia – nos recuerda Nietzsche –nunca es solo dolor: es la forma estética del conflicto, aquella que hace visible lo oscuro y, por lo tanto, lo hace pensable.

Así también para nosotros.

Aceptar el desafío de la inteligencia artificial significa aceptar el conflicto que nos obliga a repensarnos a nosotros mismos: no para borrar lo que fuimos, sino para reafirmar lo que verdaderamente vale, con más conciencia, con más rigor, con más coraje.

La inteligencia artificial, en la producción científica, puede convertirse en una ocasión de regeneración. Pero esa regeneración no vendrá por sí sola. Dependerá de nosotros. De nuestra responsabilidad al afrontar el riesgo. De nuestra medida. De nuestra libertad.

«algorithms can never predict complex events and phenomena due to the lack of adequate data. […] Even algorithms, therefore, no matter how much they increase our knowledge about the preferences and expectations of individuals, clash with the unpredictable nature of social dynamics, with the imponderable complexity of interaction between different subjects».

AI LAW

INTERNATIONAL REVIEW
OF ARTIFICIAL INTELLIGENCE LAW

Judiciary, Artificial Intelligence and Human Rights

Antonella D'Andrea

Professoressa associata di Diritto del lavoro
nell'Università degli Studi di Roma Tor Vergata

Abstract

(EN): The paper analyzes the impact of artificial intelligence (AI) on judicial activity, examining opportunities and critical issues arising from the use of algorithmic systems in decision-making processes. After reconstructing the theoretical origins of predictive justice and the reasons for the growing interest in tools capable of reducing legal uncertainty and the judicial backlog, the article explores the main applications of AI in judicial systems – from case law research to alternative dispute resolution, from risk assessment to assisted drafting of documents – while highlighting the structural limitations of algorithms. Through the analysis of emblematic cases, the risks associated with the opacity of the models, discriminatory biases and the possible compromise of the principles of due process, transparency and equality of arms are highlighted. The article argues for the need to ensure the quality and accessibility of data, the verifiability of decision-making procedures and the preservation of the autonomy of the judge, to avoid forms of technological determinism and jurisprudential conformism. The final part examines the regulatory framework, in particular the role of the GDPR and new European regulatory initiatives, outlining prospects for an AI-responsible government that ensures respect for fundamental rights and the coherence of the judicial system.

(ES): El trabajo analiza el impacto de la inteligencia artificial (IA) en la actividad judicial, examinando las oportunidades y las cuestiones críticas derivadas del uso de sistemas algorítmicos en los procesos de toma de decisiones. Tras reconstruir los orígenes teóricos de la justicia predictiva y las razones del creciente interés por herramientas capaces de reducir la incertidumbre jurídica y la sobrecarga de los órganos judiciales, el artículo explora las principales aplicaciones de la IA en los sistemas judiciales – desde la búsqueda de jurisprudencia hasta los mecanismos de resolución alternativa de conflictos, desde la evaluación del riesgo hasta la redacción asistida de documentos –, poniendo de relieve al mismo tiempo las limitaciones estructurales de los algoritmos. Mediante el análisis de casos emblemáticos, se destacan los riesgos asociados a la opacidad de los modelos, los sesgos discriminatorios y la posible afectación de los principios del debido proceso, la transparencia y la igualdad de armas. El artículo sostiene la necesidad de garantizar la calidad y la accesibilidad de los datos, la verificabilidad de los procedimientos decisorios y la preservación de la autonomía del juez, para evitar formas de determinismo tecnológico y de conformismo jurisprudencial. La parte final examina el marco regulatorio, en particular el papel del RGPD y las nuevas iniciativas regulatorias europeas, esbozando perspectivas para un gobierno

responsable de la IA que garantice el respeto de los derechos fundamentales y la coherencia del sistema judicial.

(FR): L'article analyse l'impact de l'intelligence artificielle (IA) sur l'activité judiciaire, en examinant les opportunités et les enjeux critiques découlant de l'utilisation de systèmes algorithmiques dans les processus de prise de décision. Après avoir reconstitué les origines théoriques de la justice prédictive et les raisons de l'intérêt croissant pour des outils capables de réduire l'incertitude juridique et l'engorgement des juridictions, l'article explore les principales applications de l'IA dans les systèmes judiciaires – de la recherche de jurisprudence aux mécanismes de règlement alternatif des différends, de l'évaluation des risques à la rédaction assistée de documents – tout en mettant en évidence les limites structurelles des algorithmes. À travers l'analyse de cas emblématiques, sont mis en lumière les risques liés à l'opacité des modèles, aux biais discriminatoires et à l'atteinte possible aux principes du procès équitable, de la transparence et de l'égalité des armes. L'article soutient la nécessité de garantir la qualité et l'accessibilité des données, la vérifiabilité des procédures décisionnelles et la préservation de l'autonomie du juge, afin d'éviter des formes de déterminisme technologique et de conformisme jurisprudentiel. La dernière partie examine le cadre réglementaire, en particulier le rôle du RGPD et les nouvelles initiatives réglementaires européennes, en esquissant des perspectives pour une gouvernance responsable de l'IA garantissant le respect des droits fondamentaux et la cohérence du système judiciaire.

(DE): Der Beitrag analysiert die Auswirkungen der Künstlichen Intelligenz (KI) auf die richterliche Tätigkeit und untersucht die Chancen sowie die kritischen Herausforderungen, die sich aus dem Einsatz algorithmischer Systeme in Entscheidungsprozessen ergeben. Nach einer Rekonstruktion der theoretischen Ursprünge der prädiktiven Justiz und der Gründe für das wachsende Interesse an Instrumenten, die geeignet sind, Rechtsunsicherheit und die Überlastung der Gerichte zu verringern, beleuchtet der Artikel die wichtigsten Anwendungsfelder der KI in Justizsystemen – von der Rechtsprechungsrecherche über alternative Streitbeilegungsmechanismen bis hin zur Risikobewertung und zur unterstützten Erstellung von Dokumenten – und hebt zugleich die strukturellen Grenzen von Algorithmen hervor. Anhand der Analyse exemplarischer Fälle werden die Risiken aufgezeigt, die mit der Intransparenz der Modelle, diskriminierenden Verzerrungen (Bias) sowie einer möglichen Beeinträchtigung der Grundsätze des fairen Verfahrens, der Transparenz und der Waffengleichheit verbunden sind. Der Beitrag betont die Notwendigkeit, die Qualität und Zugänglichkeit der Daten, die Überprüfbarkeit der Entscheidungsverfahren sowie die Wahrung der richterlichen Unabhängigkeit sicherzustellen, um Formen technologischen Determinismus und jurisprudenziellen Konformismus zu vermeiden. Der letzte Teil untersucht den Regulierungsrahmen, insbesondere die Rolle der DSGVO und neuer europäischer Regulierungsinitiativen, und skizziert Perspektiven für eine verantwortungsvolle *Governance* der KI, die die Achtung der Grundrechte und die Kohärenz des Justizsystems gewährleistet.

(PT): O artigo analisa os impactos da Inteligência Artificial (IA) na atividade judicial, examinando as oportunidades e os desafios críticos decorrentes do uso de sistemas algorítmicos nos processos de tomada de decisão. Após reconstruir as origens teóricas da justiça preditiva e as razões do crescente interesse por instrumentos capazes de reduzir a incerteza jurídica e a sobrecarga dos tribunais, o artigo explora os principais campos de aplicação da IA nos sistemas judiciais – desde a pesquisa de jurisprudência até os mecanismos de resolução alternativa de conflitos, desde a avaliação de riscos até a redação assistida de documentos –, ao mesmo tempo em que destaca as limitações estruturais dos algoritmos. Por meio da análise de casos exemplares, são evidenciados os riscos associados à opacidade dos modelos, aos vieses discriminatórios e à possível violação dos princípios do devido processo legal, da transparência e da igualdade de armas. O artigo enfatiza a necessidade de garantir a qualidade e a acessibilidade dos dados, a verificabilidade dos

procedimentos decisórios e a preservação da independência do juiz, a fim de evitar formas de determinismo tecnológico e de conformismo jurisprudencial. A parte final examina o marco regulatório, em particular o papel do RGPD e as novas iniciativas regulatórias europeias, delineando perspetivas para uma governança responsável da IA que assegure o respeito aos direitos fundamentais e a coerência do sistema judicial.

(IT): L'articolo analizza l'impatto dell'Intelligenza Artificiale (IA) sull'attività giudiziaria, esaminando le opportunità e le sfide critiche derivanti dall'uso di sistemi algoritmici nei processi decisionali. Dopo aver ricostruito le origini teoriche della giustizia predittiva e le ragioni del crescente interesse per strumenti capaci di ridurre l'incertezza giuridica e il sovraccarico dei tribunali, l'articolo esplora i principali ambiti di applicazione dell'IA nei sistemi giudiziari – dalla ricerca giurisprudenziale ai meccanismi di risoluzione alternativa delle controversie, dalla valutazione del rischio alla redazione assistita di documenti –, mettendo al contempo in evidenza i limiti strutturali degli algoritmi. Attraverso l'analisi di casi esemplari, vengono evidenziati i rischi associati all'opacità dei modelli, ai *bias* discriminatori e alla possibile violazione dei principi del giusto processo, della trasparenza e della parità delle armi. L'articolo sottolinea la necessità di garantire la qualità e l'accessibilità dei dati, la verificabilità delle procedure decisionali e la salvaguardia dell'indipendenza del giudice, al fine di evitare forme di determinismo tecnologico e di conformismo giurisprudenziale. La parte finale esamina il quadro regolatorio, in particolare il ruolo del RGPD e le nuove iniziative regolatorie europee, delineando prospettive per una *governance* responsabile dell'IA che assicuri il rispetto dei diritti fondamentali e la coerenza del sistema giudiziario.

Summary: 1. AI systems and judicial activity. – 2. The application of AI in judicial systems and its critical issues. – 3. The regulatory framework and new regulatory perspectives. – 4. Concluding remarks.

1. AI systems and judicial activity

Artificial intelligence (hereinafter AI) and the digitalisation of procedures now permeate every sector, including the judiciary. Beyond the development of electronic proceedings in civil, administrative and criminal matters, courts are increasingly confronted with new challenges arising from algorithmic decision-making. On the one hand, traditional legal categories must be adapted to situations in which decisions are taken by automated systems [1]. This is particularly evident in cases involving liability for harm caused by autonomous

[1] A. PAJNO-M. BASSINI-G. DE GREGORIO-M. MACCHIA-F.P. PATTI-O. POLLICINO-S. QUATTROCOLO-D. SIMEOLI-P. SIRENA, *AI: profili giuridici. Intelligenza artificiale: criticità emergenti e sfide per il giurista*, in *BioLJ*, 2019, p. 205 ff.

systems, where identifying the responsible party becomes highly problematic because established legal concepts are not easily applicable to such new circumstances [2].

On the other hand, AI is increasingly employed as a tool for predictive justice, an emerging field of law that uses algorithmic systems capable of quickly analysing enormous volumes of information to estimate the likelihood of success or failure of a case before a court [3]. The term «predictive» is, nevertheless, misleading: rather than predictions, these systems produce forecasts concerning the possible legal-logical reasoning of judges [4].

These technologies may help address judicial stagnation caused by increasing caseloads, which undermine the principle of reasonable duration of proceedings; additionally, they may strengthen legal certainty, understood as the objective foreseeability of the legal consequences attached to specific conduct [5], with

[2] U. RUFFOLO, *Intelligenza Artificiale,* machine learning *e responsabilità da algoritmo*, in *Giur. it*., 2019, p. 1689 ff.; U. RUFFOLO-E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema e in quello statunitense*, in *Giur. it*., 2019, p. 1704 ff.; R. LEENES-E. PALMERINI-B. KOOPS-A. BERTOLINI-P. SALVINI-F. LUCIVERO, *Regulatory Challenges of Robotics: Some Guidelines for Addressing Legal and Ethical Issues*, in *Law, Innov. Tech*., 9, 1, 2017, p. 1 ff.; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione europea*, in *Giur. it*., 2019, p. 1715 ff.; U. SALANITRO, *Intelligenza artificiale responsabilità: la strategia della Commissione Europea*, in *Riv. dir. civ*., 2020, p. 1246 ff.; G. DI ROSA, *Quali regole per i sistemi automatizzati intelligenti?*, *ivi*, 2021, p. 823 ff.; A. DI MARTINO, *Intelligenza artificiale e responsabilità civile in ambito sanitario*, Milan, 2022, *passim*; V. DI GREGORIO, *Intelligenza artificiale e responsabilità civile: quale paradigma per le nuove tecnologie?*, in *Danno resp.*, 2022, p. 61 ff. The case law of the Court of Justice of the European Union has addressed the application of the part of such operators for any unlawful acts committed by third-party users through the services provided (Cases C-70/10, Scarlet Extended v. SABAM e C-360/10, SABAM v. Netlog; Joined Cases C-236/8, C-237/8, C-238/08, *Google France*. Unlike decisions focusing on robotics, which have often relied on the categories of liability for defective products in relation to harmful events, a more recent line of case law seeks to reconcile the use of algorithmic systems with, on the one hand, the administration of justice and, on the other hand, the exercise of public administrative powers. The relevant rulings in this field include those of the Italian Council of State (section VI, April 8, 2019, No. 2270) and the French *Conseil Constitutionnel* (Judgment No. 2018-765 DC June 12, 2018) but also that of the Supreme Court of Wisconsin (Loomis v. Wisconsin, 881 N.W.2d 749 (Wis. 2016), certiorari denied by the Wisconsin Supreme Court, (State v. Loomis, case 2015AP17-CR, Judgment July 13th 2016, in *Harvard LR*, 130, 2017, p. 1530 ff.).

[3] L. VIOLA, *La giustizia predittiva del lavoro*, in *Lav. Dir. Eur.*, 2, 2023, p. 2 ff.; G. ARIOLLI, *Nomofilachia, giustizia predittiva e intelligenza artificiale*, in *Giust. ins.*, 3, 2023, *passim*.

[4] C. CASTELLI-D. PIANA, *Giustizia predittiva. La qualità della giustizia in due tempi*, in *Quest. Giust.*, 3, 2018, pp. 153-165; M.R. MAUGERI, *I robot e la possibile «prognosi» delle decisioni giudiziali*, in A. CARLEO (ed.), *Decisione robotica*, Bologna, 2019, p. 159 ff.; D. DALFINO, *Creatività e creazionismo, prevedibilità e predittività*, in *Foro it.*, 2018, V, c. 393 ff.; M. SCIACCA, *Algoritmo e giustizia alla ricerca di una mite predittività*, in *Pers. merc.*, 2023, p. 75.

[5] G. GOMETZ, *La certezza giuridica come prevedibilità*, Turin, 2005, *passim*; C. SALAZAR,

beneficial effects on economic and social relations and, ultimately, a deflationary effect on litigation [6].

Interest in predicting litigation outcomes can be traced back to Max Weber, who, in describing the origins of Western capitalism, highlighted the importance of legal calculation [7]. The philosophical foundation of predictive justice has also been linked to the American Legal Realist movement [8], while the idea of a «mechanical» justice dates back to eighteenth-century theories of judicial syllogism [9].

Nevertheless, the use of automated systems to support or replace judicial decision-making raises numerous concerns.

2. The application of AI in judicial systems and its critical issues

Private actors – insurance companies, law firms, and legal service providers – have driven much of the interest in AI, motivated by the desire to reduce legal uncertainty and the unpredictability of judicial decisions. Public authorities, however, increasingly face pressure to integrate these tools into services delivered to citizens. In the legal domain, AI is used for advanced case-law search engines, online dispute resolution, assistance in drafting legal documents, predictive analytics, automated contract classification and clause detection, and chatbots for informing or guiding parties [10].

A well-known example is the State v. Loomis case [11], in which the defendant

Judex ex machina*? Note su giustizia, giudici e intelligenza artificiale*, in *Consulta online*, 2021, p. 918 ff.; E. AL MUREDEN-R. ROVATTI, *Gli assegni di mantenimento tra disciplina legale e intelligenza artificiale*, Turin, 2020, *passim*; S. GOVERNATORI-M. MALTAGLIATI-G. MARLIANI-G. PACINI-V. PILLA, *Come calcolare gli assegni di mantenimento nei casi di separazione e divorzio*, Milan, 2009, *passim*.

[6] A. NATALE, *Introduzione. Una giustizia (im)prevedibile*, in *Quest. Giust*., 4, 2018, p. 7 ff.

[7] N. IRTI, *Per un dialogo sulla calcolabilità giuridica*, in A. CARLEO (ed.), *Calcolabilità giuridica*, Bologna, 2017, p. 23; ID., *Il tessitore di Goethe (per la decisione robotica)*, in A. CARLEO (ed.), *Decisione robotica*, cit., p. 17 ff.; ID., *Un diritto incalcolabile*, Turin, 2016, *passim*.

[8] G. ALPA, *Note sulla calcolabilità nel diritto nordamericano*, in A. CARLEO (ed.), *Calcolabilità giuridica*, cit., p. 88 ff.

[9] M. BARBERIS, *Giustizia predittiva: ausiliare e sostitutiva. Un approccio evolutivo*, in *Milan LR*, 3, 2, 2022, p. 8 ff.

[10] N. ALETRAS-D. TSARAPATSANIS-D. PREOȚIUC-PIETRO-V. LAMPOS, *Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective*, in *PeerJ Comp. Sc*., 2:e93, 2016.

[11] State v. Loomis, 881 N.W.2d 749, Wis. 2016. F. BASILE, *Intelligenza artificiale e diritto*

received a harsher sentence based not only on his past actions but also on his presumed future conduct, as evaluated by an algorithm used to predict recidivism risk. Although the Wisconsin Supreme Court upheld the use of the algorithm – arguing that the defendant could challenge the input data and that the sentence would not have differed absent the algorithm – it also stressed that such tools could not be the exclusive or decisive basis for judicial determinations [12]. Subsequent analysis by the NGO ProPublica, however, revealed significant racial bias, with the algorithm systematically overestimating the recidivism risk of certain racial groups [13]. These tools tend to reproduce or amplify existing social inequalities because algorithms process non-neutral data, producing deterministic outcomes that may confine individuals to group-based assessments [14].

In Italy, the debate on algorithmic decisions has often concerned automated administrative acts affecting public employees, particularly the controversial algorithm used to assign newly appointed teachers to their first workplace. The Council of State did not reject algorithmic administrative decisions in principle, recognising advantages such as reduced processing time for repetitive, non-discretionary tasks and the elimination of negligent or intentional human interference [15]. Nonetheless, it expressed caution regarding entirely automated decisions, stressing that rights protection requires human-authored, reasoned, and contestable decisions [16].

AI can certainly help overcome some judicial inefficiencies, particularly in technical assessments involving financial calculations in civil matters, such as

penale: quattro possibili percorsi di indagine, in *Dir. pen. cont.*, 10, 2019, p. 19 ff.; L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena*, *ivi*, 2, 2019, p. 362 ff.; M. BRENNER, J. SUK GERSEN, M. HALEY, M. LIN, A. MERCHANT, R. JAGDISHWAR MILLETT, S. K. SARKAR, D. WEGNER, *Constitutional Dimensions of Predictive Algorithms in Criminal Justice*, in *Harvard Civil Rights-Civil Liberties LR*, 55, 2020, p. 267 ff.

[12] State v. Loomis, 881 N.W.2d, par. 99: «a COMPAS risk assessment is only one of many factors that may be considered and weighed at sentencing».

[13] J. DRESSEL, H. FARID, *The accuracy, fairness, and limits of predicting recidivism*, in *Sc. Adv.*, 17.1.2018.

[14] F. PASQUALE, *The Black Box Society: The Secret Algorithms that Control Money and Information*, Cambridge-London, 2015, *passim*, spec. p. 102 ff. On the risks of discrimination arising from the use of algorithms: T. CALDERS, I. ŽLIOBAITĖ, *Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures*, in B. CUSTERS, T. CALDERS, B. SCHERMER, T. ZARSKY (Eds.), *Discrimination and Privacy in the Information Society. Data Mining and Profiling in Large Databases*, Berlin, Heidelberg, 2013, p. 43 ff.; A. D'ALOIA, *Il diritto verso "il mondo nuovo". Le sfide dell'Intelligenza Artificiale*, in *BioLJ*, 1, 2019, p. 16 ff.

[15] Cons. Stato, Sez. VI, 08/04/2019, n. 2270, § 8.

[16] Cons. Stato, Sez. VI, 08/04/2019, n. 2270, § 9.

compensation for dismissal, disability percentage determinations, calculation of spousal support, child maintenance, or monetary damages. AI can also support factual reconstruction, preparation of draft decisions, and handling of simple or repetitive documentary cases. It can be useful for alternative dispute resolution, for anonymisation and pseudonymisation of judicial decisions (considering Recital 26 and Art. 4(5) General Data Protection Regulation, GDPR), and for logistical support such as scheduling hearings or reserving courtrooms.

AI systems, however, present structural and technical challenges. Their performance can be compromised by flawed training data [17], or by «statistical discrimination», where outputs are influenced by pre-existing statistical correlations [18]. Algorithms are shaped by the values and intentions of their creators [19], and their results require interpretation according to inherently subjective data categorizations [20].

When algorithmic structures are protected by intellectual property or trade secrets, preventing access to their functioning, they escape user and judicial scrutiny, contradicting the principle of equality of arms [21]. A defendant must be able to examine and challenge the scientific validity of the tool used against them. Trade-secret opacity severely restricts transparency and accountability.

Safeguards must therefore ensure data quality, independence of data sources, transparency of inputs, and explainability of decision paths to prevent systemic discriminatory impacts [22]. Yet many AI systems, especially machine-learning models, function as «black boxes» [23]. Undoubtedly, similar remarks can also be made to human decisions [24], where they are adopted on the basis of unconscious

[17] F. PASQUALE, *op. cit*., p. 101 ff.

[18] P. HACKER, *Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law*, in *CMLR*, 2018, p. 1146 ff.

[19] S. SIGNORATO, *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo*, in *Riv. dir. proc.*, 2020, p. 614.

[20] G. UBERTIS, *Intelligenza artificiale e giustizia predittiva*, in *Sist. pen.*, 16 ottobre 2023, p. 5 ff.

[21] P.P. PAULESU, *Intelligenza artificiale e giustizia penale. Una lettura attraverso i principi*, in *Arch. pen. Web*, 1, 2022, 11 gennaio 2022, p. 3; M. LUCIANI, *La decisione giudiziaria robotica*, in *Riv. AIC*, 2018, p. 889.

[22] EU Agency for Fundamental Rights, *Bias in algorithms – Artificial Intelligence and Discrimination*, Luxembourg, 2022, *passim*.

[23] S. SHALEV-SHWARTZ-S. BEN-DAVID, *Understanding Machine Learning: From Theory to Algorithms*, Cambridge, 2014, *passim*.

[24] G. LEGNINI, *Introduzione*, in A. CARLEO (ed.), *Decisione robotica*, cit., p. 12; M.E. FRANKEL, *Criminal Sentences. Law Without Order*, New York, 1973, *passim*.

impulses and conditioning according to the ancient Roman adage *Habent sidera sua lites* [25]; nevertheless, this cannot justify the admissibility of algorithmic decisions that affect the legal sphere of a subject without the latter being able to check the correctness of the computer process. Concerns also arise regarding the potential influence of algorithmic recommendations on judicial impartiality [26]. Judges may avoid diverging from algorithmic forecasts to reduce personal exposure to criticism, especially where judicial mandates depend on re-election or where disciplinary or civil liability regimes are strict [27]. The «practical force» of algorithms may lead to *de facto* automatic decision-making, conferring a misleading scientific authority on algorithmic outputs [28]. This dynamic risks generating dangerous jurisprudential conformism [29], inhibiting the evolution of judicial reasoning and preventing the law from responding to societal changes. Since predictive justice operates inductively [30], overreliance on precedent-based algorithms can reinforce pre-existing biases and cement majority trends.

While expanding access to legal knowledge and improving judicial reason-

[25] L. SOLIDORO, Habent sua sidera lites*: sulla storia dell'imponderabile nell'avventura processuale*, in *TSDP*, 2018, p. 1 ff.; V. TENORE, *I robot in giudizio! considerazioni sull'utilizzo di intelligenza artificiale (ai) da parte del magistrato: abdicare al ragionamento e alla riflessione quotidiana a favore di un robot togato, ovvero un "cretino digitale", è un bene per la giustizia e per i suoi attori e destinatari?*, in *Riv. Corte Conti*, 4, 2023, p. 1 ff.

[26] D. KAHNEMAN-O. SIBONY-C.R. SUNSTEIN, *Noise: A Flaw in Human Judgment*, London, 2021, trad. it. di E. Gallitelli, *Rumore. Un difetto del ragionamento umano*, Milan, 2021, *passim*; R. RICHARDSON-J.M. SCHULTZ-K. CRAWFORD, *Dirty data, bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, in *New York University LR Online*, 94, 2019, p. 15 ff.; S. ARCIERI, *Bias cognitivi e decisione del giudice: un'indagine sperimentale*, in *Diritto Penale e Uomo*, 2 aprile 2019, p. 1 ff.; M. SOZIO, *La decisione giudiziale in condizioni di incertezza: tra distorsioni cognitive e neuroscienze*, in *Quad. Dip. Ion.*, 11, 2019, p. 355 ff.

[27] A. SIMONCINI-S. SUWEIS, *Samar, Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Riv. fil. dir.*, 8, 1, 2019, p. 100; C. BARBARO, *Uso dell'intelligenza artificiale nei sistemi giudiziari: verso la definizione di principi etici condivisi a livello europeo?*, in *Quest. Giust.*, 4, 2018, p. 194; P. COMOGLIO, *Prefazione*, in J. NIEVA-FENOLL, *Intelligenza artificiale e processo*, Turin, 2019, pp. X-XI.

[28] A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLJ*, 1, 2019, p. 81.

[29] P. ZELLINI, *La dittatura del calcolo*, Milan, 2018, p. 15; V. TENORE, *op. cit.*, p. 26; A. GARAPON-J. LASSÈGUE, *La giustizia digitale. Determinismo tecnologico e libertà*, Bologna, 2021, *passim*; P.P. PAULESU, *op. cit.*, p. 23; A. SANTOSUOSSO-G. SARTOR, *La giustizia predittiva: una visione realistica*, in *Giur. it.*, 2022, p. 1761; L. DE RENZIS, *Primi passi nel mondo della giustizia «high tech»: la decisione in un corpo a corpo virtuale fra tecnologia e umanità*, in A. CARLEO (ed.), *Decisione robotica*, cit., p. 141.

[30] M. BARBERIS, *op. cit.*, p. 6; V. TENORE, *op. cit.*, pp. 3 e 26.

ing is desirable, automated systems that constrain judicial autonomy and independence pose constitutional risks and facilitate deterministic or quantitative decision models [31].

The need for transparency in algorithmic decision-making and for fairness in computerized decision-making processes must be guaranteed across all sectors and in every situation in which choices are made that affect fundamental rights, freedoms, and legitimate interests of individuals, as required by the provisions of the General Data Protection Regulation (GDPR) No. 2016/679, which, pursuant to Article 5, also applies when data are processed within AI systems. Article 15 GDPR, in particular, establishes that the data subject has the right to know of the existence of automated decision-making, including profiling, and, in such cases, to receive meaningful information about the logic involved as well as the significance and the envisaged consequences of such processing for the data subject [32]. Based on this provision, it has been argued that the data controller must also provide information regarding the existence of potential discriminatory risks inherent in the algorithm used [33].

Article 57 GDPR grants national supervisory authorities significant investigative, verification, and enforcement powers, while Article 83 provides for considerable administrative fines in cases of violation. Articles 35 et seq. GDPR further require a specific data protection impact assessment when a particular type of processing – especially involving new technologies – may present a high risk to the rights and freedoms of individuals (Recital 71 GDPR).

Finally, Article 22(1) GDPR allows the data subject to object to decisions that produce significant effects on their legal sphere and are based solely on

---

[31] M. SCIACCA, *Algoritmo e giustizia alla ricerca di una mite predittività*, in *Dir. Giust. Cost.*, 12 maggio 2023, p. 9. A. CARRATTA, *Decisione robotica e valori del processo*, in *Riv. dir. proc.*, 2020, p. 99, states that positive law emerges from the case-by-case application of legal provisions by judges, who interpret and apply them in the light of general principles and system-derived values. Judicial decision-making frequently involves the use of open-textured concepts and general clauses, as well as a non-neutral process of reconstructing and selecting the legally relevant facts. This process is not a mechanical or arithmetical operation, but rather a value-laden assessment that entails an additional element of unpredictability, particularly at the fact-finding stage, where evidentiary materials are evaluated. On this issue, see also M.B. MAGRO, *Human Decision-Making and Robotic Decision-Making. A Hypothesis of Liability for Robotic Procreation*, in *Legisl. pen.*, 10 May 2020, *passim*.

[32] F. PIZZETTI, *La protezione dei dati personali e la sfida dell'intelligenza artificiale*, in F. PIZZETTI (ed.), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Turin, 2018, p. 30 ff.

[33] P. HACKER, *op. cit.*, p. 1174; E. PELLECCHIA, *Profilazione e decisioni automatizzate al tempo della* black box society*: qualità dei dati e leggibilità dell'algoritmo nella cornice della* responsible research and innovation, in *Nuove leggi civ. comm.*, 2018, p. 1224 ff.

automated processing, including profiling. These are decisions in which no human involvement in the decision-making process is present, meaning no involvement that goes beyond a merely symbolic act [34]. However, paragraph 2 provides exceptions where the decision: (a) is necessary for entering into or performing a contract between the data subject and the controller; (b) is authorised by Union or Member State law to which the controller is subject, provided that the law also sets out suitable measures to safeguard the data subject's rights, freedoms, and legitimate interests; or (c) is based on the data subject's explicit consent. In the cases referred to in paragraph 2(a) and (c), the controller must nonetheless implement appropriate measures to safeguard the data subject's rights, freedoms, and legitimate interests, and the data subject retains the right to obtain human intervention by the controller, to express their point of view, and to contest the decision.

The GDPR expressly excludes from its scope the processing of data for the purposes of the prevention, investigation, detection, or prosecution of criminal offences or the execution of criminal penalties. Therefore, the provisions mentioned may be invoked by anyone with an interest in verifying the fairness of the procedure followed by an AI system. This firm obligation to ensure transparency in algorithmic decision-making has also been reaffirmed in the guidelines on the use of AI published by the European Commission [35].

Case law and decisions issued by courts also show that the extraordinary variety of concrete cases a judge must resolve cannot be reduced to the abstract schemes of algorithmic computation, as these cases often require complex choices that cannot be predetermined [36].

The use of AI as a substitute for the judge therefore affects guarantees inherent in jurisdiction, such as: the effectiveness and completeness of the parties' right to a defence; the quality of judicial decision-making; the judge's ability to bring to light the irreducible particularity of facts; the judge's ability to calibrate the decision to those facts; and the obligation to provide reasoning.

These principles are enshrined not only in several national constitutions but also within the European legal order. Title V of the Treaty on the Functioning of the European Union (Articles 67-89) provides for the creation of an area of

[34] Data Protection Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation (EU) 2016/679* (wp251rev.01), adopted on 3 October 2017 as last Revised and Adopted on 6 February 2018, p. 20 ff.

[35] European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *Building Trust in Human-Centric Artificial Intelligence Communication*, Brussels, 8.4.2019, COM(2019) 168 final.

[36] C. CASTELLI-D. PIANA, *Giusto processo e intelligenza artificiale*, Santarcangelo di Romagna, 2019, pp. 86-87.

freedom, security, and justice in which the application of the Charter of Fundamental Rights of the European Union plays a crucial role. Within this area, access to justice is a fundamental right, enshrined both in Article 47 of the Charter – which guarantees the right to an effective remedy and to an impartial tribunal – and in Article 6 of the European Convention on Human Rights, signed by the EU and its Member States.

For these reasons, EU institutions have long debated the use of new information technologies in the service of justice, particularly from a rights-protective and human-rights-compliant perspective [37], with the aim of supporting «a high level of data protection, digital rights and ethical standards» [38].

Since the publication of the study Liability for artificial intelligence and other emerging digital technologies in April 2019, and the adoption of the White Paper on AI of 19 February 2020 by the European Commission [39], the opportunities associated with the use of new technologies and the risks they pose to fundamental rights have been highlighted, creating the need for a new regulatory framework. In the subsequent conclusions of 21 October 2020, the Council called for addressing the opacity, complexity, bias, unpredictability, and partially autonomous behaviour of certain AI systems in order to ensure compatibility with fundamental rights and facilitate the application of legal rules [40].

The European Parliament has also undertaken several initiatives. In October 2020, it adopted a set of resolutions concerning AI, including with respect to ethics, liability, and copyright [41]. In 2021, additional resolutions followed on AI in criminal matters (2020/2016(INI)) and in the fields of education, culture, and the audiovisual sector (2020/2017(INI)). The Council of the European Union subsequently invited the Commission to seize the opportunities of digitalisation while ensuring that fundamental rights are preserved in all EU actions aimed at establishing effective legal protection and access to justice [42].

---

[37] M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Dir. pen. cont.*, 29 maggio 2019, p. 12.

[38] European Council, *Conclusions* EUCO 14/17, Brussels, 19 October 2017, p. 7.

[39] European Commission, *White Paper on Artificial Intelligence – A European approach to excellence and trust*, Brussels, 19.2.2020, COM(2020) 65 final, *passim*, spec. p. 9 ff.

[40] Council of the European Union, Presidency conclusions – The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change, Brussels, 21 October 2020, 11481/2020, p. 5.

[41] European Parliament Resolution, *Intellectual property rights for the development of AI technologies*, 20 October 2020, 2020/2015(INI).

[42] Council of the European Union, Council Conclusions on the Use of Artificial Intelligence in the Field of Justice, Brussels, 16 December 2024, 16933/24.

Within this framework lies Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (AI Act), which establishes rules on AI, based on the conviction that its use may contribute to socially and environmentally beneficial outcomes and provide essential competitive advantages to businesses and the European economy (European Commission, COM(2021) 206 final). However, recognising that the same elements and techniques that generate socioeconomic benefits may also entail risks or negative consequences for individuals and society, the EU has committed to a balanced approach, ensuring technological leadership while protecting EU values, fundamental rights, and principles.

The AI Act classifies AI systems used in the legislative, justice, and alternative dispute resolution sectors as high-risk and subjects them to conformity assessments, quality certification requirements for data and transparency, and other safeguards aimed at ensuring a high level of reliability (AI Act Annex III, para. 8(a)).

Conversely, several practices are prohibited because they are considered incompatible with fundamental rights and EU values: real-time facial recognition (except for terrorism-related offences); the use of predictive policing to project past statistical patterns into the future; emotion-recognition techniques; social scoring systems that classify citizens based on social behaviour or subjective characteristics.

Moreover, data collection and open access to databases must not allow the profiling of judges to reconstruct their legal reasoning or judicial approach with the aim of building predictive defence strategies.

AI must therefore be used in justice systems in full compliance with the Charter of Fundamental Rights, the AI Act, the GDPR, and Directive (EU) 2016/680 regarding the processing of personal data by competent authorities for criminal purposes, ensuring transparency, accountability, and oversight, and guaranteeing that AI systems are safe and trustworthy. Key criteria in the design, development, and use of AI systems must include: the right to an effective remedy, the right to a fair and public hearing within a reasonable time, and the presumption of innocence.

In any case, AI may support decision-making in judicial and alternative dispute resolution proceedings, but must never replace the final decision-making, which must remain a human-led activity. The use of AI in this context must be transparent, properly traceable, and explainable, and subject to human supervision.

In light of these considerations, the European e-Justice Strategy (2024–2028) aims to guide the digital transformation of the justice sector throughout the EU by offering concrete and lasting benefits through reduced costs associ-

ated with access to justice and with the functioning of judicial systems. In particular, the Strategy aims to: identify strategic and operational objectives and principles to be upheld during the digital transformation process; implement organisational and methodological measures; identify key factors that facilitate and promote digitalisation; and promote mechanisms facilitating coordination and follow-up of progress made in these initiatives. These objectives must be pursued in parallel in both civil and criminal justice, taking into account the implications of digital transformation in these areas.

Currently, regulations on the service of documents (Regulation (EU) 2020/1784) and the taking of evidence (Regulation (EU) 2020/1783) already establish a legal framework for the digitalisation of these instruments of judicial cooperation in civil and commercial matters. The regulation on electronic evidence (Regulation (EU) 2023/1543) is an additional instrument for digitalisation in criminal proceedings. The regulation and directive on the digitalisation of cross-border judicial cooperation and access to justice (the so-called «digitalisation package») are fundamental pillars for European e-Justice in the coming years.

Beyond the EU, several other international organisations have adopted significant documents in the field of AI, including: the OECD Council Recommendation on Artificial Intelligence; the UN Human Rights Council Resolution of 10 July 2024 on the promotion and protection of all human rights – including independence and impartiality of the judiciary, jurors, assessors, and independence of lawyers; the final report of the UN Advisory Body on Artificial Intelligence; and the Council of Europe Framework Convention on AI, human rights, democracy, and the rule of law [43], which aims to ensure that AI activities remain fully consistent with human rights, democracy, and the rule of law.

3. The regulatory framework and new regulatory perspectives

Of particular importance is the European Ethical Charter on the Use of AI in Judicial Systems and its guidelines on the use of AI in the judiciary, issued by the European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe.

It is a non-binding instrument, yet it has the merit of setting out both substantive and methodological principles. It is particularly significant insofar as it constitutes the first document aimed at establishing criteria to guide the development and use of AI systems supporting judicial decision-making. The under-

[43] Council of Europe Treaty Series No. 225, Vilnius, 5 September 2024.

lying idea emerging from the document is that AI – when used not as a substitute for the judge but as an auxiliary tool – may, under certain circumstances, enhance predictability in the application of the law and improve the uniformity of judicial orientations. The fundamental principles governing the use of AI in judicial systems are, however, the respect for fundamental rights and the right to non-discrimination, the right to system quality and safety, and the right to transparency and user oversight. Compliance with these principles must be ensured from the very stages of system design and learning, according to an ethical-by-design or human-rights-by-design approach. The observance of such principles should be guaranteed by an independent authority entrusted with certification and supervisory powers.

The Charter also examines specific issues concerning analytical methodology and the automated processing of judicial decisions. Data processing conducted through machine learning should rely on original certified documents, and the integrity of such data should be ensured throughout all phases of processing. Particular attention should be devoted to the selection of judicial decisions that are subsequently processed through machine learning. These decisions should be genuinely representative of the various realities on which judges are called to rule, and should not correspond to predetermined analytical grids. Moreover, the Charter emphasises the need for a secure technological environment for the storage and implementation of machine-learning models and algorithms. It therefore underscores the importance of making data-processing techniques accessible and comprehensible, as well as permitting external audits by independent actors in order to detect potential distortions. Finally, it highlights the need to strengthen users' autonomy when employing AI tools and services. Judges, in particular, should be able at any time to return to the judicial decisions and data used to produce a given output, and must always retain the possibility of departing from such outputs in light of the specific circumstances of the case at hand [44]. Each user should also be informed, in clear and comprehensible language, of the binding or non-binding nature of AI-generated solutions, of the available options, and of his or her right to legal assistance and to access a court.

4. Concluding remarks

Every individual has the right to «an effective remedy by the competent national tribunals for acts violating the fundamental rights granted him by

[44] G. Riccio, *Ragionando su intelligenza artificiale e processo penale*, in *Arch. pen*., 3, 2019, p. 10.

the constitution or by law» (Art. 8 of the Universal Declaration of Human Rights).

The advancement of digitalisation therefore requires an assessment of the effectiveness of the protection afforded to fundamental rights and to the core principles of due process: the presumption of innocence; the adversarial principle and the ability to challenge the evidence; the duty to give reasons for judicial decisions and the control of their legality and rationality; and the principle *in dubio pro reo.*

Moreover, digital-justice initiatives must respect the independence of the judiciary and the rule of law. Particular caution is needed in the field of criminal justice, where the use of technologies for remote communication may entail serious risks for the fundamental rights of suspects and defendants – especially the right to a fair trial, the right to be present at one's trial, and defence rights. Indeed, every individual «is entitled in full equality to a fair and public hearing by an independent and impartial tribunal, in the determination of his rights and obligations and of any criminal charge against him» (Art. 10 UDHR), and if accused of a criminal offence «has the right to be presumed innocent until proved guilty according to law in a public trial at which he has had all the guarantees necessary for his defence» (Art. 11 UDHR).

It must not be forgotten that every trial is a tool for the protection of individual rights, and that behind each request for judicial protection «there lies a personal struggle», such that «logic and statistics may assist, but prove inadequate when confronted with the human story underlying the proceedings» [45]. Accordingly, the limit that cannot be crossed is the safeguarding of the human context of justice, preserving its capacity for mediation and conflict resolution.

As with any phenomenon influenced by technological evolution, it is crucial to pay careful attention both to the potential risks (for the system of guarantees and individual protections, as well as for the exercise of judicial functions) and to the opportunities (cost and time savings relating to data-collection operations, large-scale document analysis, research into judicial precedents, or the drafting of standardised acts).

AI technologies play an increasingly decisive and pervasive role in social activities geared toward maximising functional efficiency, including the legal sector. Yet digital innovation will yield positive results only if firmly anchored to the goals of broadening access to remedies for fundamental human rights violations, facilitating the understanding of judicial reasoning, reducing inconsistency and shortening trial duration, and – more broadly – improving the or-

[45] F. DALFINO, *Stupidità (non solo) artificiale, predittività e processo*, in *Quest. Giust.*, 3 luglio 2019.

ganisation of the judicial machinery. In this context, the training of all judicial actors is indispensable.

A cautious approach to integrating AI into judicial systems must therefore be recommended: it will become increasingly important to ensure that proposed applications genuinely add value for users and effectively contribute to access to justice and to the achievement of the objectives of the public justice service, rather than undermining the inviolable rights of individuals. Above all, judicial supervision must remain central and must constitute an indispensable step because human qualities can never be replaced by artificial components and «the healthy awareness of one's own ignorance is a feature that characterises human beings and their search for meaning, and does not lend itself to the logic of AI» [46].

[46] C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *DPCE*, Special Issue, May, 2019, p. 124.

AI LAW

INTERNATIONAL REVIEW
OF ARTIFICIAL INTELLIGENCE LAW

L'*A.I.* e il suo possibile supporto nel processo di valutazione del nesso di causalità nel giudizio di colpevolezza

Nicola Scerbo

Dottore di ricerca in Ordine giuridico ed economico europeo
presso l'Università Magna Graecia di Catanzaro

Abstract

(EN): The implementation process of recent artificial intelligence systems is also represented by the considerations of professionals regarding the possibility that A.I. models. generative methods can improve the cognitive process in the mechanisms of assessing guilt. Considering the recent regulatory innovations in the unitary and internal context, driven by the need to provide a regulatory barrier to the proliferation of algorithms and the related risks involved, it is possible to hypothesize future scenarios of implementation of the technologies in question in the judicial field. The objective is to articulate a reasoning regarding the possible implementation of AI systems. in the penal system, without neglecting the enormous critical issues connected, but also evaluating the possible positive elements to support the process of evaluating the causal links placed under scrutiny by the judge.

(ES): El proceso de implementación de los recientes sistemas de inteligencia artificial también se refleja en las consideraciones de los profesionales acerca de la posibilidad de que los modelos de IA y los métodos generativos puedan mejorar el proceso cognitivo en los mecanismos de valoración de la culpabilidad. Teniendo en cuenta las recientes innovaciones regulatorias, tanto en el contexto unitario como en el interno, impulsadas por la necesidad de establecer una barrera normativa frente a la proliferación de algoritmos y los riesgos conexos, es posible hipotetizar escenarios futuros de implementación de las tecnologías en cuestión en el ámbito judicial. El objetivo es articular un razonamiento sobre la posible implementación de sistemas de IA en el sistema penal, sin descuidar las enormes cuestiones críticas asociadas, pero evaluando también los posibles elementos positivos que puedan apoyar el proceso de valoración de los nexos causales sometidos al escrutinio del juez.

(FR): Le processus de mise en œuvre des récents systèmes d'intelligence artificielle se reflète également dans les considérations des professionnels quant à la possibilité que les modèles d'IA et les méthodes génératives puissent améliorer le processus cognitif dans les mécanismes d'évaluation de la culpabilité. Compte tenu des récentes innovations réglementaires, tant dans le contexte unitaire que dans le contexte interne, motivées par la nécessité d'établir une barrière nor-

mative face à la prolifération des algorithmes et aux risques connexes, il est possible d'hypothétiser des scénarios futurs de mise en œuvre des technologies en question dans le domaine judiciaire. L'objectif est d'articuler un raisonnement sur la possible mise en œuvre de systèmes d'IA dans le système pénal, sans négliger les importantes questions critiques qui y sont liées, tout en évaluant également les éléments positifs susceptibles de soutenir le processus d'évaluation des liens causaux soumis à l'examen du juge.

(DE): Der Implementierungsprozess der jüngsten Systeme der künstlichen Intelligenz spiegelt sich auch in den Überlegungen der Fachleute wider, wonach KI-Modelle und generative Methoden den kognitiven Prozess in den Mechanismen der Schuldbeurteilung verbessern könnten. In Anbetracht der jüngsten regulatorischen Neuerungen, sowohl im einheitlichen als auch im innerstaatlichen Kontext, die durch die Notwendigkeit motiviert sind, eine normative Schranke gegenüber der Verbreitung von Algorithmen und den damit verbundenen Risiken zu errichten, ist es möglich, zukünftige Szenarien der Implementierung der betreffenden Technologien im justiziellen Bereich zu hypothetisieren. Ziel ist es, eine Argumentation zur möglichen Implementierung von KI-Systemen im Strafrechtssystem zu entwickeln, ohne die damit verbundenen erheblichen kritischen Fragen zu vernachlässigen, zugleich aber auch die möglichen positiven Elemente zu bewerten, die den Prozess der Bewertung der kausalen Zusammenhänge unterstützen können, die der Prüfung durch den Richter unterliegen.

(PT): O processo de implementação dos sistemas mais recentes de inteligência artificial reflete-se também nas considerações dos profissionais quanto à possibilidade de que modelos de IA e métodos generativos possam melhorar o processo cognitivo nos mecanismos de avaliação da culpabilidade. Tendo em conta as recentes inovações regulatórias, tanto no contexto unitário quanto no interno, motivadas pela necessidade de estabelecer uma barreira normativa frente à proliferação de algoritmos e aos riscos a ela associados, é possível hipotetizar cenários futuros de implementação das tecnologias em questão no âmbito judicial. O objetivo é articular um raciocínio sobre a possível implementação de sistemas de IA no sistema penal, sem negligenciar as importantes questões críticas a ela associadas, mas também avaliando os possíveis elementos positivos que possam apoiar o processo de avaliação dos nexos causais submetidos ao escrutínio do juiz.

(IT): Il processo di implementazione dei più recenti sistemi di intelligenza artificiale si riflette anche nelle considerazioni dei professionisti circa la possibilità che modelli di IA e metodi generativi possano migliorare il processo cognitivo nei meccanismi di valutazione della colpevolezza. Tenendo conto delle recenti innovazioni regolatorie, tanto nel contesto unitario quanto in quello interno, motivate dalla necessità di stabilire una barriera normativa di fronte alla proliferazione degli algoritmi e ai rischi ad essa associati, è possibile ipotizzare scenari futuri di implementazione delle tecnologie in questione in ambito giudiziario. L'obiettivo è articolare un ragionamento sulla possibile implementazione di sistemi di IA nel sistema penale, senza trascurare le rilevanti questioni critiche ad essa connesse, ma valutando anche i possibili elementi positivi che possano sostenere il processo di valutazione dei nessi causali sottoposti allo scrutinio del giudice.

Sommario: 1. Premessa. – 2. I sistemi di *A.I.* e lo stato dell'arte. – 3. Il ragionamento sul potenziale delle moderne *A.I.* nella giustizia penale. – 4. Le novità normative unitarie ed interne. – 5. Quali prospettive per il diritto penale nella fase cognitiva. – 6. Il giudizio di colpevolezza ed il nesso di causalità (leggi scientifiche di copertura). – 7. Il possibile ruolo delle *A.I.* nell'accertamento del nesso causale. – 8. Conclusioni.

1. Premessa

Elemento caratterizzante il processo di implementazione dei recenti sistemi di intelligenza artificiale presso la società odierna è rappresentato dal continuo ragionamento intorno ai potenziali impieghi della potente tecnologia nei vari contesti dell'agire umano, soprattutto quale supporto collaborativo ai fini dell'efficientamento o arricchimento dei processi produttivi [1].

La domanda che ci si pone ai fini del ragionamento giuridico è, soprattutto, quanto i recenti sviluppi in merito ai modelli di *A.I.* generativi possano migliorare il processo cognitivo dei meccanismi di valutazione della colpevolezza, in particolar modo in ambito penale, viste le notevoli ripercussioni sugli aspetti più sensibili della natura umana [2].

Da sempre, quando di discute di colpevolezza dell'individuo ci si muove su un terreno scivoloso, i cui contorni possono subire erosioni o addizioni speculative in funzione del giudizio promosso o del particolare fatto oggetto di valutazione.

Nondimeno risulterebbe quantomai aleatorio discutere di colpevolezza senza entrare nei profili teorici e pratici di un sistema di valori quantomai abusato e frutto di evoluzioni giurisprudenziali anche audaci, ai limiti del rispetto della tipicità penale [3].

---

[1] V. l'importante lavoro di McKinsey & Company, *The economic potential of generative AI: The next productivity frontier*, 14 giugno 2023, e quello di International Data Corporation (IDC), *Drive $1 Trillion in Productivity Gains by 2026,* 12 dicembre 2023. Per approfondimenti v. anche la Rilevazione tecnologica *L'Intelligenza Artificiale in banca: stato dell'arte e prospettive* – Anno 2020 – Rapporto conclusivo pubblicata sul sito del CIPA. Come specificato sul sito della banca d'Italia la «*Rilevazione sull'IT nel settore bancario italiano – Profili tecnologici e di sicurezza*, curata annualmente dalla CIPA (Convenzione Interbancaria per l'Automazione) in collaborazione con l'ABI, fornisce un quadro aggiornato sull'utilizzo dell'*Information and Communication Technology* nel settore bancario nazionale, con particolare riguardo alle scelte *IT* in materia di metodologie e tecnologie innovative a supporto dell'operatività bancaria».

[2] V. B. Galgani, *Forme e garanzie nel prisma dell'innovazione tecnologica. Alla ricerca di un processo penale "virtuoso"*, Milano, 2022, p. IX, richiamando le opere caricaturali di Honoré Daumier raffiguranti *les gens de justice*.

[3] Sul tema, v. A. Provera, *La 'causalità impossibile'. La probabilità logica senza una legge*

Tralasciando gli aspetti dei comportamenti di natura dolosa, rimessi ad altre analisi e studi di diversa natura, ci si vuole soffermare in particolar modo sul processo di valutazione della colpevolezza e sulla elefantiaca problematica nel giudizio di causalità, dinanzi al quale molti giuristi e addetti ai lavori hanno trovato non poche ostilità interpretative, dovute essenzialmente a difetti ontologici del contesto di riferimento [4].

Il rispetto del nucleo fondamentale dei diritti umani è consustanziale al processo di implementazione di qualsiasi tecnologia di supporto al sistema di valutazione processuale, anche in virtù del radicato senso di umanesimo antropocentrico nelle scienze giuridiche contemporanee.

È, pertanto, indispensabile muoversi con cautela e rigore logico onde evitare facili speculazioni o ragionamenti estemporanei e poco utili al miglioramento complessivo della discussione in divenire, circa il possibile impiego dei sistemi di intelligenza artificiale a supporto del processo di valutazione della probabilità statistica nel ragionamento controfattuale, posto a carico del giudice in fase di valutazione del nesso di causalità [5].

In un articolato reticolo di innovazioni tecnologiche annunciate, quotidianamente, dagli addetti ai lavori, il più delle volte per ragioni di carattere speculativo e commerciale, è altresì ragionevole approcciare il fenomeno con le necessarie riserve, dovute in particolar modo al non eludibile pensiero di immanente facilitazione dell'agire nella complessa macchina giudiziaria, quantomai farraginosa e sovraccarica di oneri accessori difficilmente sovrapponibili e conciliabili con le necessità di evasione ordinarie per le istante sottoposte al vaglio cognitivo del giudice [6].

Muovendo dai principi teorico-pratici delle odierne tecnologie di intelligenza artificiale, alla luce delle recenti innovazioni normative in ambito unitario

scientifica di copertura", in *SP*, 7-8, 2023, p. 5 ss. E ancora, sempre sul tema della causalità, v. V. PACILEO, *Ancora sul tema della causalità nel giudizio penale. Le ragioni del diritto*, *ivi*, 4, 2024, p. 5 ss.

[4] La sentenza "Franzese", d'altro canto, pur avendo introdotto i primi passi verso la valorizzazione del metodo induttivo, ha richiesto ed esplicitato la necessaria presenza di una legge scientifica di copertura, nonostante la stessa possa avere probabilità statistica *bassa*, qualora dal compendio probatorio emerga comunque una spiegazione causale convincente e «la sicura non incidenza nel caso di specie di altri fattori interagenti in via alternativa» (c.d. *alta probabilità logica*). Per approfondimenti si rinvia a Cass., sez. un., 11 luglio 2022, n. 30328.

[5] Sul problema della causalità come "origine" di tutte le altre difficoltà applicative del diritto penale nella materia del danno da prodotto cfr. C. PIERGALLINI, *Danno da prodotto e responsabilità penale. Profili dommatici e politico-criminali*, Milano, 2004.

[6] Per approfondimenti sui temi legati ai tempi della giustizia in Italia, si rinvia alla pagina dedicata al «monitoraggio della giustizia civile e penale» sul sito del ministero della giustizia (www.giustizia.it).

ed interno, mosse dalla necessità, secondo il legislatore di turno [7], di arrecare un argine regolatorio al proliferare degli innumerevoli algoritmi ed ai relativi rischi connessi ad un possibile uso improprio, sarà possibile ipotizzare eventuali scenari futuri di implementazione delle tecnologie in questione in ambito giudiziario, alcuni dei quali forse già non troppo inverosimili o in corso di sperimentazione presso ordinamenti giuridici non troppo distanti dal nostro per valori di riferimento e cultura.

Allo stesso tempo non può mancare un esame attento della teoria generale circa il giudizio di colpevolezza, nel tentativo di articolare un ragionamento in merito alla possibile implementazione dei sistemi di *A.I.* nel sistema penale, nonché sulle evidenti ed enormi criticità connesse, senza, però, tralasciare i possibili elementi di positività del possibile impiego dei sistemi di *A.I.* a supporto del processo di valutazione coerentemente con quanto prospettato dal legislatore nei suoi recentissimi interventi [8].

2. I sistemi di *A.I.* e lo stato dell'arte

Non tutti sono edotti del fatto che i moderni sistemi di intelligenza artificiale non sono poi così diversi da alcuni algoritmi già in uso da anni o addirittura decenni.

Il processo alla base del funzionamento è essenzialmente caratterizzato dalla capacità di analisi di dati in *input* non dissimile da sistemi di calcolo cui eravamo abituati fin dagli albori dell'informatica, scaturenti *output* in ragione delle necessità e degli obiettivi di volta in volta richiesti all'operato della macchina [9].

Ciò che realmente è cambiato, nella gestione delle moderne tecnologie di intelligenza artificiale, è l'enorme capacità computazionale, resa possibile dai continui progressi in ambito informatico e dalle tecnologie *hardware* poste al servizio degli algoritmi concepiti in software di elaborazione dei dati.

Non meno importante per portata e rischio gestionale è la spropositata mole di dati cui è possibile accedere, anche grazie al contributo quotidiano degli

[7] V. in tal senso M. LIBERTINI-M.R. MAUGERI-E. VINCENTI, *Intelligenza artificiale e giurisdizione ordinaria una ricognizione delle esperienze in corso*, in A. PAJNO-F. DONATI-A. PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, II, *Amministrazione, responsabilità, giurisdizione*, Bologna, 2023, p. 483 ss.

[8] V. M. RAMAJOLI, *Le questioni sull'intelligenza artificiale ancora da affrontare sono molte: il diritto è solo uno dei tanti strumenti per cercare di risolverle*, in *Il Mulino*, 25 marzo 2023.

[9] Per tale motivo secondo la definizione comune, almeno dagli anni Cinquanta, «l'*A.I.* è un qualsiasi sistema ingegnerizzato che risolve problemi con successo, in modo tale che se dovessimo farlo noi dovremmo essere intelligenti».

utenti in rete, agenti più o meno consapevoli dell'enorme massa di informazione cui i sistemi di intelligenza artificiale può accedere per la sua formazione od addestramento, in un continuo processo di automiglioramento ed affinamento del responso erogato [10].

L'architettura odierna del word wide web, essenzialmente proiettata al management open source ha senza ombra di dubbio facilitato il processo di implementazione costante delle tecnologie di *A.I.*, da cui, a ragion veduta sono scaturite le prime iniziative di confinamento e gestione del rischio da parte del legislatore unitario, interno e d'oltreoceano [11].

Nondimeno, anche sistemi ordinamentali meno avvezzi alla condivisione dei dati o essenzialmente strutturati in forme di protezionismo autarchico, sono esenti dal rischio di una cattiva gestione nello sviluppo ed implementazione dei moderni sistemi di intelligenza artificiale, ma, per ovvie ragioni di trasparenza burocratica amministrativa, sarà difficile carpirne le criticità e le prospettive di accesso ai sistemi di *A.I.* odierni.

Non è, altresì possibile, per tali ordinamenti, vista l'impossibilità di analizzare in modo genuino dati esperienziali di applicazione dei sistemi in questione, esprimere giudizi calibrati ed approfonditi in merito.

Sono noti, a titolo di esempio, alcuni fatti di cronaca relativi ad implementazioni nei processi di valutazione cognitiva in alcune provincie cinesi, con sostituzione completa dell'agire umano, in particolar modo della figura del giudice, nonché relativa sottoposizione a giudizio esclusivo della macchina e relativo sistema di calcolo [12].

---

[10] Si pensi, a titolo di esempio, al tema legato alla privacy, o più propriamente, alla tutela del trattamento dei dati personali, in considerazione della gran mole di dati che questi sistemi di *A.I.* (a volte equipaggiati per l'appunto di sensori e telecamere avanzate) possono acquisire in relazione alla vita, anche privata, dei cittadini coinvolti dal loro utilizzo nei più svariati ambiti. Situazioni in cui i dati oggetto del processo di elaborazione dell'algoritmo potrebbero essere in varie forme manipolati, sottratti o deformati, anche con notevole pregiudizio per le persone cui essi fanno rifermento. Sul tema v. F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto penale e uomo, rivista online,* 29 settembre 2019.

[11] Come primi accenni di intervento sul tema v. S. QUATTROCOLO, *Intelligenza artificiale e giustizia: nella cornice della Carta etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *www.lalegislazionepenale.eu*, 18 dicembre 2018. V. anche tra gli altri G. CANZIO, *Il dubbio e la legge*, in *Dir. pen. cont.*, 2018, p. 1 ss.; M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assesment tools tra Stati Unite ed Europa*, *ivi*, 2019, p. 1 ss.; A. NATALE, *Introduzione. Una giustizia (im) prevedibile?*, in *Quest. Giust.*, 4, 2018, p. 1 ss.; nello stesso fascicolo, v. pure i contributi di C. COSTANZI, *La matematica del processo: oltre le colonne d'Ercole della giustizia penale* e di C. CASTELLI-D. PIANA, *Giustizia predittiva. La qualità della giustizia in due tempi*.

[12] L'esempio tipico è quello del procuratore *robot* nell'ambito del processo penale per la va-

Difficile averne la certezza, ma è ragionevole ipotizzare altri scenari simili in altrettanti contesti poco inclini alla dimensione dei valori democratici.

Dal canto nostro, non mancano esempi più o meno sperimentali di implementazione dei sistemi di *A.I.* in ambito giudiziario, il più delle volte, però, con notevoli elementi critici e, soprattutto, la presenza di una supervisione umana posta ad argine delle possibili e plausibili distorsioni agli equilibri nel sistema delle garanzie processuali [13].

Le diverse possibili interpretazioni poste al vaglio del giudice non dipendono solo dai differenti valori di riferimento, ma, soprattutto, dalla travagliata, complessa ed incerta attività dell'interprete in un momento ed in un assetto come l'attuale in cui è presente una realtà multi-fonte in continua e rapida evoluzione che impone una sapiente ed articolata ricostruzione dell'ordinamento e della gerarchia delle fonti [14].

---

lutazione delle prove, dei presupposti per custodia cautelare e arresto e della pericolosità dell'indagato/imputato. Il sistema, sviluppato da un'equipe di ricercatori sotto la guida del Prof. Shi Young utilizza per la valutazione un *database* di 17000 casi avvenuti tra il 2015 ed il 2020, ma, al momento, è in grado di analizzare solo otto fattispecie di reato previste dal Codice penale cinese (nella fattispecie frodi con carte di credito, gioco d'azzardo, guida pericolosa, lesioni internazionali, intralcio ai doveri d'ufficio, furto, frode e "*scelta di litigi e provocazione di guai*"). Secondo quanto dichiarato dal Prof. Young, dai primi *test* emergerebbe un margine di errore del 3% nella valutazione del grado colpevolezza od innocenza dell'imputato, ma senza che il *robot* possa prendere parte alla parte attiva della decisione. Lo stesso interverrebbe come mero ausilio al libero convincimento del giudice. Infatti, sempre secondo il prof. Young, affinché la macchina possa prendere autonomamente decisioni avrebbe bisogno di «convertire un linguaggio umano complesso ed in continua evoluzione in un formato matematico o geometrico standard che un computer potrebbe capire». Per ulteriori approfondimenti si veda l'articolo pubblicato sul *South China Morning Post* il 26 dicembre 2021 al seguente link: *https://www.scmp.com*.

[13] È il caso di impiego del cosiddetto sistema COMPAS, acronimo di *Correctional offender management profiling for alternative sanctions*: algoritmo coperto da brevetto e sviluppato da un'azienda privata di nome *Equivant*, presentato come strumento di valutazione della possibilità di recidiva di un imputato sulla base di una serie di risposte ad un questionario standard di 137 quesiti. Si tratta di uno dei tanti sistemi di valutazione del rischio a disposizione dell'apparato giudiziario statunitense, impiegato per ricevere informazioni utili a valutare la possibilità di reati futuri e le tipologie di misure da applicare nella prevenzione. I primi accenni della giurisprudenza italiana in tal senso provengono dal settore del diritto amministrativo. Una recente sentenza del Cons. St., sez. VI, 8 aprile 2019, n. 2270, ha legittimato l'impiego di algoritmi nelle procedure valutative dalla pubblica amministrazione, a condizione, però, che vi sia trasparenza e possibilità di controllo giurisdizionale. I giudici hanno precisato che l'impiego di algoritmi e procedure automatizzate è considerato un «atto amministrativo informatico» e, come tale, deve necessariamente sottostare ai principi generali di ragionevolezza, pubblicità, proporzionalità e trasparenza. Per ulteriori approfondimenti, v. E. CHITI-B. MARCHETTI-N. RANGONE, *L'impiego di sistemi di intelligenza artificiale nelle pubbliche amministrazioni italiane: prove generali*, in A. PAJNO-F. DONATI-A. PERRUCCI (a cura di), *op. cit*., p. 43 ss.

[14] Per approfondimenti v. C. CASTELLI, *Giustizia predittiva*, in *Quest. Giust.*, 8/2/2022.

3. Il ragionamento sul potenziale delle moderne *A.I.* nella giustizia penale

L'impiego di moderni strumenti di elaborazione computazionale di autoapprendimento, altrimenti detta *A.I.* (*Artificial Intelligence*), si basa sulla costruzione di specifici algoritmi (*software*) in grado di alimentare un sistema progressivo ed automatizzato di ricerca delle innumerevoli possibili correlazioni con modelli di riferimento agglomerati in appositi contenitori di dati, stoccati in strutture fisiche di memoria (*hardware*).

La velocità disarmante con la quale una simile tecnologia sta progressivamente interessando ogni aspetto del vivere quotidiano, ha parallelamente iniziato a coinvolgere settori più sensibili, quali la giustizia, spingendo studiosi e, non solo, a chiedersi quali possano essere le possibili implicazioni nell'impiego di simili strumenti in dinamiche afferenti alla tutela delle più svariate libertà individuali [15].

Non mancano esempi pratici nell'impiego di sistemi sperimentali di intelligenza artificiale in ambito penale, anche se, le esperienze più significative al riguardo, come già accennato, provengono per il momento da realtà non perfettamente in linea con il sistema di garanzie della realtà europea [16].

L'esperienza già maturata nell'impiego dell'*A.I.* ha mostrato evidenti criticità d'impiego efferenti ad enormi *bias* di pregiudizio su parametri di genere o razziali ed errori ricorrenti nell'approccio valutativo ad eventi radicalmente innovativi, caratterizzati da un gradiente di ragionamento laterale difficilmente raggiungibile da apparati tecnici preimpostati su modelli di riferimento standardizzati e schemi di associazione lineare [17].

---

[15] Al riguardo v. G.M. BACCARI-M. MARRAFFINO, *Le prospettive di utilizzo delle chatbot nel procedimento penale*, in *Dir. pen. proc.*, 2021, p. 1008 ss.; E. NAGNI, *Artificial intelligence, l'innovativo rapporto di (in)compatibilità fra machina* sapiens *e processo penale*, in *SP*, 7, 2021, p. 5 ss.; F. DELVECCHIO, *L'informatizzazione della giustizia penale*, in *Dir. pen. cont.*, 2, 2021, p. 60 ss.; R.E. KOSTORIS, *Predizione decisoria,* diversion *processuale e archiviazione*, in *SP*, 23/7/2021; A. LAVORGNA-G. SUFFIA, *La nuova proposta europea per regolamentare i Sistemi di Intelligenza Artificiale e la sua rilevanza nell'ambito della giustizia penale*, in *Dir. pen. cont.*, 2, 2021, p. 88 ss. I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in *Riv. it. dir. proc. pen.*, 2021, p. 83 ss.; ID., *Il diritto penale dei software a "duplice uso"*, in R. WENIN-G. FORNASARI (a cura di), *Diritto penale e modernità. Le nuove sfide fra terrorismo, sviluppo tecnologico e garanzie fondamentali*, Napoli, 2017, p. 361 ss.

[16] La commissione europea per l'efficienza della giustizia ha elaborato nel 2018 la *Carta etica per l'impiego dell'intelligenza artificiale nei sistemi giudiziari ed ambiti connessi.* Per ulteriori approfondimenti si rinvia al seguente link: https://rm.coe.int.

[17] Il più classico degli esempi impiegato per enfatizzare la fallibilità dei sistemi di Intelligenza artificiale o altrimenti detto di autoapprendimento (*machine learning*), è il programma di

Nel caso *Loomis* contro *Wisconsin*, in cui fu impiegato il sistema COMPAS nella valutazione del rischio di recidiva dell'imputato, la Corte Suprema, nella consapevolezza del potenziale rischio di impiego della tecnologia in questione, caratterizzata dalla possibilità di generare un pregiudizio di genere, ha comunque ritenuto legittimo il suo impiego in funzione supplementare al legittimo convincimento del giudice chiamato a valutare il caso concreto.

Uno dei principi stabiliti, poi divenuto cardine anche nelle varie proposte di regolamentazione dell'*A.I.* anche in ambito europeo, è il costante e consapevole controllo umano su tale tecnologia, la quale mai potrà sostituirsi interamente alla libera ed autonoma valutazione del giudicante, ma, semmai, potrà essere impiegata alla stregua di un supporto facilitante nella gestione del processo di elaborazione del dato documentale e casistico [18].

Non c'è dubbio che simili sistemi, dato il loro enorme potenziale in termini di razionalizzazione delle risorse umane e fisiche, parallelamente all'inevitabile e progressiva evoluzione tecnologica, entreranno sicuramente a far parte dell'arsenale strumentale messo a disposizione del sistema giudiziario, nell'ambito di

gestione della guida autonoma per autoveicoli. Nonostante gli enormi investimenti nel settore ed i notevoli progressi tecnologici, il sistema in questione ha costantemente mostrato limiti significati in quelli che con ricorrenza sistematica possono essere identificati come dilemmi etici o paradossi della scelta. Data una condizione ineludibile di potenziale danno a sé stessi o ad altra cosa o persona, evento del tutto possibile in caso di incidentalità stradale, ci si è sempre chiesti come potesse reagire una macchina nella scelta tra arrecare un danno a un guidatore o ad un passante, scegliendo di andare a sbattere per evitare una conseguenza al posto dell'altra (il c.d. «paradosso del treno» nelle varie formulazioni). Per approfondimenti v. P. FOOT, *The Problem of Abortion and the Doctrine of the Double Effect in Virtues and Vices*, in *Oxford R*, 5, 1967, p. 1 ss.; v. anche P. LIN, *The Ethics of Autonomous Cars*, in *The Atlantic*, 8/10/2013; EMERGING TECHNOLOGY FROM THE ARXIV, *Why Self-Driving Cars Must Be Programmed to Kill*, in *MIT TR*, 22/10/2015. Il problema delle situazioni "dilemmatiche" ha suscitato particolare interesse in dottrina, v. al riguardo A. ENGLANDER, *Das selbstfahrende Kraftfahrzeug und die Bewältigung dilemmatischer Situationen*, in *ZIS*, 2016, p. 609 ss.; A. HEVELKE-J. NINDA-RUMELIN, *Selbstfahrende Autos und Trolley-Probleme: Zum Aufrechnen von Menschenleben im Falle unausweichlicher Unfälle*, in *JWE*, 19, 2015, p. 5 ss.; E. HILGENDORF, *Autonomes Fahren im Dilemma. Überlegungen zur moralischen und rechtlichen Behandlung von selbsttätige Kollisionsvermeidesystemen*, in E. HILGENDORF (hrsg), *Autonome Systeme und neue Mobilität*, Baden-Baden, 2017, p. 143; J.C. JOERDEN, *Zum Einsatz von Algorithmen in Notstandslagen. Das Notstandsdilemma bei selbstfahrenden Kraftfahrzeugen als strafrechtliches Grundlagenproblem*, *ivi*, p. 73 ss.; G.M. SANDER-J. HOLLERING, *Strafrechtliche Verantwortlichkeit im Zusammenhang mit automatisiertem Fahren*, in *NStZ*, 2017, p. 193 ss.; C. ROXIN-L. GRECO, *Strafrecht. Allgemeiner Teil*, I, München, 2020, Rn. 125d ss.; I. COCA VILA, *Self-driving Cars in Dilemmatic Situations: An Approach Based on the Theory of Justification in Criminal Law*, in *CLP*, 1, 2018, p. 59 ss.

[18] Per ulteriori approfondimenti v. H.-W. LIU-C.-F. LIN, Y.-J. CHEN, *Beyond State v Loomis – artificial intelligence, government algorithmization and accountability*, in *IJLIT*, 27, 2019, p. 122 ss.

attività giudicanti, corredando ulteriormente il già variegato prontuario di attrezzi del mestiere, potenzialmente a rischio di abusi ed uso improprio nella più elementare gestione delle libertà individuali.

Le opzioni tecnologiche disponibili alimentano il timore circa la possibile visone meccanicistica del ruolo del giudice, edulcorato o, se si vuole, mascherato dall'impiego delle nuove tecniche nella prospettiva di decisione automatizzate ed indifferenti alla realtà umana.

Si tratta di una prospettiva molto inquietante, alla quale si potrebbe dare risposta concreta alimentando il grado di consapevolezza sull'esistenza di possibili fattori esterni ed interni, dal potenziale condizionante in merito a decisioni sensibili, nel tentativo di sterilizzarne gli effetti principali attraverso l'uso di una discrezionalità ragionata, motivata e trasparente.

In pratica, non una vera e propria sostituzione del giudica con la macchina, ma un contributo che può essere erogato dall'intelligenza artificiale alla sua già innata capacità di elaborazione dei dati, ma velocizzandone notevolmente i tempi.

Da tali considerazioni emerge come i sostenitori di un uso sostitutivo dell'*A.I.* mediante decisioni automatizzate, sull'assunto di possibili risvolti positivi in termini di oggettività, non suscettibilità a pressioni esterne, ma, soprattutto, tempistiche più brevi, non si rendano conto delle possibili conseguenze di una simile determinazione.

Come già accennato, le decisioni delle varie *A.I.* derivano dalla qualità ed integrità dei dati impiegati nel processo elaborativo, sia negli aspetti processuali che nei precedenti giurisprudenziali, pertanto, elemento essenziale nell'implementazione di tali sistemi, dovrebbe essere la costante ed agevole controllabilità dei parametri anzidetti [19].

D'altro canto, alcuni elementi di riflessione su eventuali pregi meritano un breve cenno, infatti, in generale, ciò che maggiormente caratterizza il processo decisionale umano non è solo riferibile ad elementi di valutazione interni al processo come le prove, le condizioni delle parti, eventuali precedenti ecc., ma anche a caratteri esterni del tutto riconducibili alla personalità del giudice ed avvenimenti riguardanti la vita dello stesso con coinvolgimento emotivo.

Alcuni esempi, tutti afferenti al settore penale, pongono molti interrogativi. In alcuni studi si specifica che i principali vantaggi nell'affidarsi ad algoritmi,

[19] Sul tema v. M. LANZI, Self-driving cars *e responsabilità penale. La gestione del "rischio stradale" nell'era dell'intelligenza artificiale*, Torino, 2023, pp. 172 ss., 198 ss., favorevole rispetto ad una limitazione alla sola "colpa grave"; v. anche A. FIORELLA, *Responsabilità penale del Tutor e dominabilità dell'Intelligenza Artificiale. Rischio permesso e limiti di autonomia dell'Intelligenza Artificiale*, in R. GIORDANO-A. PANZAROLA-A. POLICE-S. PREZIOSI-M. PROTO (a cura di), *Il diritto nell'era digitale. Persona, Mercato, Amministrazione, Giustizia*, Milano, 2022, p. 660 ss.

piuttosto che ad esseri umani, non risiede tanto nella superiorità delle regole quanto nella loro assenza di rumore [20] (cosiddetta dispersione casuale).

Si può affermare che allo stato non esiste una linea univoca sull'impiego di nuove tecnologie di intelligenza artificiale nella giustizia penale, ma è più ragionevole ipotizzare un approccio progressivo a quella che può essere considerata una possibile rivoluzione copernicana nella gestione del comparto giustizia.

La chiave di volta, probabilmente, risiede in un atteggiamento pacato ed esplorativo di un fenomeno che, per quanto avversato dalla dottrina e da parti consistenti degli addetti ai lavori, può offrire alcune opportunità meritevoli di considerazione, alla luce dei potenziali rischi di una tecnologia lasciata a sé stessa o, peggio ancora, a mani poco esperte.

È fondamentale comprenderne in modo oculato i meccanismi di fondo, senza magnificarne il potenziale e, come per ogni altra precedente evoluzione della tecnica, mitigarne le criticità.

4. Le novità normative unitarie ed interne

I recenti progressi tecnologici nel contesto internazionale hanno sollecitato notevolmente l'intervento del legislatore, fortemente preoccupato delle possibili ripercussioni dell'implementazione dei moderni sistemi di *A.I.* nei vari aspetti dell'agire umano.

In particolar modo, a livello euro unitario è subito emersa l'esigenza di un intervento regolatorio, in chiave sia definitoria dei sistemi interessati, che delimitativa del possibile impiego nei diversi contesti dell'agire umano [21]. A dir la

[20] V. D. KAHNEMAN-O. SIBONY-C.R, SUNSTEIN, *Rumore. Un difetto del ragionamento umano*, trad. E. Gallitelli, Milano, 2021, p. 15.

[21] L'art. 22 *General Data Protection Regulation* (GDPR), recepito in ambito nazionale dall'art. 8 d.lgs. n. 51/2018, aveva già introdotto il divieto di decisioni basate unicamente su trattamenti automatizzati. Infatti, sono «1. Sono vietate le decisioni basate unicamente su un trattamento automatizzato, compresa la profilazione, che producono effetti negativi nei confronti dell'interessato, salvo che siano autorizzate dal diritto dell'Unione europea o da specifiche disposizioni di legge. 2. Le disposizioni di legge devono prevedere garanzie adeguate per i diritti e le libertà dell'interessato. In ogni caso è garantito il diritto di ottenere l'intervento umano da parte del titolare del trattamento. 3. Le decisioni di cui al comma 1 non possono basarsi sulle categorie particolari di dati personali di cui all'articolo 9 del regolamento UE, salvo che siano in vigore misure adeguate a salvaguardia dei diritti, delle libertà e dei legittimi interessi dell'interessato. 4. Fermo il divieto di cui all'articolo 21 della Carta dei diritti fondamentali dell'Unione europea, è vietata la profilazione finalizzata alla discriminazione di persone fisiche sulla base di categorie particolari di dati personali di cui all'articolo 9 del regolamento UE».

verità, già in passato si era prospettata una tale esigenza [22], ma vuoi per la reticenza fisiologica delle burocrazie sovranazionali in alcuni ambiti, vuoi per lo scetticismo verso annunci più o meno immaginifici di innovazioni mirabolanti e rivoluzionarie [23], si è tergiversato in attesa di una migliore comprensione del fenomeno, o, più probabile, per mera pigrizia normativa.

Ai tempi, non era evidentemente chiaro il potenziale di sistemi in grado di stravolgere da un giorno all'altro la percezione creativa ed il ruolo superiore dell'essere umano nelle dimensione terrena [24], tuttavia, nel momento in cui è esploso senza soluzione di causa il tormentone delle *A.I.* generative, quali *ChatGPT* o simili, non è stato più possibile esimersi dalla pianificazione di interventi normativi d'impatto, volti a delimitare i confini di una innovazione tecnologica quantomeno inquietante sotto molteplici punti di vista [25].

Il rischio di uno stravolgimento improvviso dei paradigmi consolidati di interpretazione e percezione di ciò che è reale e ciò che invece può essere generato artificialmente da una macchina o, meglio, da un algoritmo di elaborazione dei dati, ha reso più evidente un bisogno primario di protezione e tutela di valori consolidati e raramente messi in discussione.

È necessario soffermare l'attenzione sull'*A.I. act*, primo vero corpo normativo con effetti prescrittivi generalmente applicabili in riferimento alla tematica in oggetto.

Una prima definizione segna in modo chiaro ed antropocentrico l'obiettivo vincolante dell'impiego dell'intelligenza artificiale ovvero «un sistema automatizzato progettato per funzionare con diversi livelli di autonomia e che può mostrare capacità di adattamento dopo l'installazione e che, per obiettivi espliciti o impliciti, deduce, dagli *input* che riceve, come generare *output* quali pre-

[22] Per un'attenta analisi della *Carta etica europea* v. S. QUATTROCOLO, *Intelligenza artificiale e giustizia: nella cornice della Carta etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *Legisl. pen*., 18 dicembre 2018, p. 4.

[23] In tal senso v. G. GIGERENZER, *Perché l'intelligenza umana batte ancora gli algoritmi*, Milano, 2023, p. 13, per il quale «restare intelligenti significa comprendere le potenzialità e i rischi delle tecnologie digitali, ed essere determinati a mantenere il controllo in un mondo popolato da algoritmi».

[24] La differenza ontologica che persiste tra umani e *A.I.* è riconosciuta da molti: E.J. LARSON, *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*, Cambridge, 2021, *passim*; N. CRISTIANINI, *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, Bologna, 2023, *passim*; M. CHIRIATTI, *Incoscienza artificiale. Come fanno le macchine a prevedere per noi*, Roma, 2021, *passim.*

[25] Come noto, *ChatGPT* è un modello di *A.I.* "generativo", liberamente accessibile online, che, attraverso algoritmi di apprendimento automatico, genera risposte agli *input* scritti forniti dall'utente. Già in diversi casi è stata riscontrata la comunicazione di notizie false da parte del sistema.

visioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali».

È molto importante concentrare l'attenzione sull'uso delle parole *imput* ed *output,* in quanto strettamente legate all'agire umano, coerentemente col sistema valoriale europeo, fondato su libertà e responsabilità nel rispetto dei principi enunciati dalla Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi di giustizia penale e nei relativi ambienti, emanata dalla Commissione europea per l'efficienza della giustizia (CEPEJ) [26].

Il ragionamento giuridico alla base dell'*A.I. act* deve necessariamente tradursi in interventi di tipo nazionale ed allo stesso tempo il formante giurisprudenziale chiamato a pronunciarsi prima dei relativi interventi normativi non potrà esimersi dall'implementazione dei sistemi di intelligenza artificiale nel processo rispettosi dei principi stabiliti in ambito europeo.

L'architettura complessiva dell'*A.I. act* non rende possibile la sostituzione del giudice con l'intelligenza artificiale e, tuttavia, rende verosimile la possibilità di un supporto della medesima tecnologia. Basato essenzialmente su una gradazione del rischio in funzione della tipologia di *A.I.* impiegata, il regolamento europeo, non prevedendo un divieto assoluto sull'uso dell'intelligenza artificiale in ambito giurisdizionale, probabilmente, ne stimola un utilizzo conforme alla rinnovata società democratica europea nel rispetto di alti *standard* qualitativi e professionali degli attori coinvolti consapevole delle possibili ricadute negative per i diritti [27].

Riguardo al tema del possibile impiego di sistemi autonomi decisionali già il GDPR riserva già nell'art. 22 uno specifico dettato normativo relativo al divieto di essere sottoposti a decisioni interamente automatizzate: così al comma 1 si specifica che «l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona» [28].

---

[26] La CEPEJ (*European Commission for The Efficiency of Justice*) è un organismo del Consiglio d'Europa composto da tecnici, rappresentativo dei 47 Stati che ne fanno parte. Il suo obiettivo è monitorare l'efficienza e la qualità dei sistemi giudiziari europei.

[27] Così B. GALGANI, *Considerazioni sui "precedenti" dell'imputato e del giudice al cospetto dell'A.I. nel processo penale*, in *SP*, 4, 2020, p. 87; EAD., *Giudizio penale,* habeas data *e garanzie fondamentali*, in *Arch. pen*., 1, 2019, pp. 23-24, ove si sottolinea con decisione che l'algoritmo non potrà mai diventare un "dittatore", ma semmai «fungere da indicatore strumentale per orientare il libero convincimento del giudice» che, tralasciando per ragioni evidenti i comodi eccessi dell'arbitrio, non si può irrigidire «in una sorta di "matematizzazione" dei criteri ermeneutici».

[28] Così, ai considerando collegati 71 e 72 «l'interessato dovrebbe avere il diritto di non essere sottoposto a una decisione, che possa includere una misura, che valuti aspetti personali che lo

L'imputato, secondo tale criterio, avrebbe sempre il diritto di ricevere informazione sull'esistenza di un processo decisionale automatizzato e sui meccanismi alla base di tale funzionamento, incluse le potenziali conseguenze giuridiche del suo impiego.

La copertura del GDPR garantirebbe l'informativa circa l'esistenza di un processo decisionale automatizzato [29], sulla logica alla base del suo funzionamento e, soprattutto, sulle conseguenze derivanti dal suo impiego.

riguardano, che sia basata unicamente su un trattamento automatizzato e che produca effetti giuridici che lo riguardano o incida in modo analogo significativamente sulla sua persona, quali il rifiuto automatico di una domanda di credito online o pratiche di assunzione elettronica senza interventi umani. Tale trattamento comprende la «profilazione», che consiste in una forma di trattamento automatizzato dei dati personali che valuta aspetti personali concernenti una persona fisica, in particolare al fine di analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze o gli interessi personali, l'affidabilità o il comportamento, l'ubicazione o gli spostamenti dell'interessato, ove ciò produca effetti giuridici che la riguardano o incida in modo analogo significativamente sulla sua persona. Tuttavia, è opportuno che sia consentito adottare decisioni sulla base di tale trattamento, compresa la profilazione, se ciò è espressamente previsto dal diritto dell'Unione o degli Stati membri cui è soggetto il titolare del trattamento, anche a fini di monitoraggio e prevenzione delle frodi e dell'evasione fiscale secondo i regolamenti, le norme e le raccomandazioni delle istituzioni dell'Unione o degli organismi nazionali di vigilanza e a garanzia della sicurezza e dell'affidabilità di un servizio fornito dal titolare del trattamento, o se è necessario per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento, o se l'interessato ha espresso il proprio consenso esplicito. In ogni caso, tale trattamento dovrebbe essere subordinato a garanzie adeguate, che dovrebbero comprendere la specifica informazione all'interessato e il diritto di ottenere l'intervento umano, di esprimere la propria opinione, di ottenere una spiegazione della decisione conseguita dopo tale valutazione e di contestare la decisione. Tale misura non dovrebbe riguardare un minore. Al fine di garantire un trattamento corretto e trasparente nel rispetto dell'interessato, tenendo in considerazione le circostanze e il contesto specifici in cui i dati personali sono trattati, è opportuno che il titolare del trattamento utilizzi procedure matematiche o statistiche appropriate per la profilazione, metta in atto misure tecniche e organizzative adeguate al fine di garantire, in particolare, che siano rettificati i fattori che comportano inesattezze dei dati e sia minimizzato il rischio di errori e al fine di garantire la sicurezza dei dati personali secondo una modalità che tenga conto dei potenziali rischi esistenti per gli interessi e i diritti dell'interessato e che impedisca tra l'altro effetti discriminatori nei confronti di persone fisiche sulla base della razza o dell'origine etnica, delle opinioni politiche, della religione o delle convinzioni personali, dell'appartenenza sindacale, dello *status* genetico, dello stato di salute o dell'orientamento sessuale, ovvero che comportano misure aventi tali effetti. Il processo decisionale automatizzato e la profilazione basati su categorie particolari di dati personali dovrebbero essere consentiti solo a determinate condizioni. La profilazione è soggetta alle norme del presente regolamento che disciplinano il trattamento dei dati personali, quali le basi giuridiche del trattamento o i principi di protezione dei dati. Il comitato europeo per la protezione dei dati istituito dal presente regolamento («comitato») dovrebbe poter emanare orientamenti in tale contesto».

[29] I primi accenna della giurisprudenza italiana in tal senso provengono dal settore del diritto amministrativo. Una recente sentenza del Consiglio di Stato, la n. 2270 dell'8 aprile 2019, sez. VI, ha legittimato l'impiego di algoritmi nelle procedure valutative dalla pubblica amministrazione, a

Le coperture del segreto industriale, indicano, allo stesso tempo, un limite di accesso alle formule matematiche e computazionali impiegate dall'algoritmo, ma deve comunque avere accesso a tutta una serie di informazione che consentano una comprensione semplice ed intelligibile del suo funzionamento [30].

Ne deriva una serie di deroghe specifiche, coperte da riserva di legge, in base a cui può essere limitata la portata degli obblighi derivanti dal regolamento, a condizione di rispettare i parametri essenziali delle *libertà fondamentali* attraverso misure *necessarie e proporzionate* in una *società democratica.*

Il quadro normativo non perfettamente agevole, unito alla fervente accelerazione del fenomeno legato all'implementazione delle *A.I.,* ha spinto l'unione, per il tramite della commissione ad iniziare un percorso di regolamentazione specifica sul tema, nell'ottica dell'armonizzazione delle regole e del contenimento degli gli aspetti critici legati all'uso della tecnologia di *A.I*., soprattutto con riguardo alla tutela delle libertà fondamentali, senza, però, trascurarne i potenziali elementi di carattere positivo [31].

condizione, però, che vi sia trasparenza e possibilità di controllo giurisdizionale. I giudici hanno precisato che l'impiego di algoritmi e procedure automatizzate è considerato un *atto amministrativo informatico* e, come tale, deve necessariamente sottostare ai principi generali di ragionevolezza, pubblicità, proporzionalità e trasparenza.

[30] Il Parlamento europeo, nel *report* del 2019 *A.I. and Robotics,* ha specificato la necessità di avere la cosiddetta *intellligibility of decisions*, nonché il diritto della persona interessata dal trattamento dei dati di essere informato sulla logica del meccanismo di elaborazione e sulla presenza di un controllo umano esterno.

[31] Nell'*annual report lays out the challenges of protecting fundamental rights in the digital age del 10 Dicembre 2021* pubblicato dalla Commissione dell'Unione europea annualmente per una verifica sul rispetto della Carta dei diritti nell'UE è specificato che «the increasing use of artificial intelligence systems can yield great benefits, but certain applications are complex and opaque, which can be a challenge for compliance with or enforcement of fundamental rights. Many Member States have developed national strategies on artificial intelligence to ensure transparency, traceability and robustness and find effective ways to comply with fundamental rights. In April 2021, the Commission proposed a legislative act to ensure that artificial intelligence systems that pose a high-risk to fundamental rights are appropriately tested and documented». Per approfondimenti, si veda il *report* sul sito ufficiale della commissione europea. Come specificato al paragrafo della relazione 1.1 alla proposta di regolamentazione «l'uso dell'intelligenza artificiale, garantendo un miglioramento delle previsioni, l'ottimizzazione delle operazioni e dell'assegnazione delle risorse e la personalizzazione dell'erogazione di servizi, può contribuire al conseguimento di risultati vantaggiosi dal punto di vista sociale e ambientale nonché fornire vantaggi competitivi fondamentali alle imprese e all'economia europea. Tale azione è particolarmente necessaria in settori ad alto impatto, tra i quali figurano quelli dei cambiamenti climatici, dell'ambiente e della sanità, il settore pubblico, la finanza, la mobilità, gli affari interni e l'agricoltura. Tuttavia, gli stessi elementi e le stesse tecniche che alimentano i benefici socioeconomici dell'*A.I.* possono altresì comportare nuovi rischi o conseguenze negative per le persone fisiche o la società».

Secondo la Commissione, una mitigazione dei rischi connessi all'uso dell'*A.I.* può avvenire mediante una legislazione complementare con norme proporzionate e flessibili. In tale modo, l'unione europea dovrebbe divenire un riferimento *gold standard* globale nel settore dell'intelligenza artificiale.

Infatti, come specificato dalla Commissione, dal quadro giuridico emerge che sviluppatori ed utenti della tecnologia avrebbero un quadro giuridico preciso basato su quattro diversi livelli di rischio: inaccettabile, elevato, limitato e minimo.

Alla luce del notevole progresso tecnologico, l'unione europea, a detta della commissione, con il *A.I. act*, si sarebbe dotata di uno strumento equilibrato in grado di preservare «la leadership tecnologica dell'*UE* e assicurare che i cittadini europei possano beneficiare di nuove tecnologie sviluppate e operanti in conformità ai valori, ai diritti fondamentali e ai principi dell'Unione» [32].

Alcune recenti conclusioni del consiglio europeo specificano, come criterio e cartina al tornasole per il rispetto dei diritti fondamentali e l'agevolazione nell'applicazione delle norme di riferimento, di tenere sempre in considerazione l'opacità, la complessità, la faziosità ed un certo grado di imprevedibilità nel comportamento di taluni sistemi di *A.I.* [33].

Come specificato anche nella relazione alla proposta di regolamento «i requisiti di qualità elevata dei dati, documentazione e tracciabilità, trasparenza, sorveglianza umana, precisione e robustezza sono strettamente necessari per attenuare i rischi per i diritti fondamentali e la sicurezza posti dall*'A.I.* e che non sono oggetto di altri quadri giuridici in vigore».

Allo stato attuale, nonostante l'approvazione definitiva dell'*A.I. act*, dopo un lungo ed estenuante percorso di trattative tra gli organi dell'unione, la fase del c.d. *trilogo* [34], dunque, non è possibile analizzare l'impatto che avrà nel tempo

[32] Come si può leggere sempre al paragrafo 1.1 della relazione alla proposta di regolamentazione con la stessa «si tiene fede all'impegno politico della presidente Von Der Leyen nei suoi orientamenti politici per la Commissione 2019-2024. L'unione europea ha annunciato che la Commissione avrebbe presentato una normativa per un approccio europeo coordinato alle implicazioni umane ed etiche dell'intelligenza artificiale. A seguito di tale annuncio la Commissione ha pubblicato il 19 febbraio 2020 il Libro bianco sull'intelligenza artificiale – Un approccio europeo all'eccellenza e alla fiducia. Il Libro bianco definisce le opzioni strategiche su come conseguire il duplice obiettivo di promuovere l'adozione dell'*A.I.* e affrontare i rischi associati a determinati utilizzi di tale tecnologia».

[33] Consiglio dell'Unione Europea, *Conclusioni della presidenza – La Carta dei diritti fondamentali nel contesto dell'intelligenza artificiale e della trasformazione digitale*, 11481/20, 2020.

[34] Nel contesto della procedura legislativa ordinaria dell'Unione europea, un trilogo è un negoziato interistituzionale informale che riunisce rappresentanti del Parlamento europeo, del Consiglio dell'Unione europea e della Commissione europea. L'obiettivo di un trilogo è raggiungere un accordo provvisorio su una proposta legislativa accettabile sia per il Parlamento che per il Consiglio, i colegislatori. Tale accordo provvisorio deve quindi essere adottato da ciascuna delle

il *corpus* normativo sull'intero processo di implementazione della nuova tecnologia nei sistemi nazionali, i quali, presumibilmente, apporteranno in seguito smussamenti ed adeguamenti specifici resi necessari da esercizi di compromesso delle varie parti coinvolte, tenendo anche conto del fatto che il regolamento entrerà completamente, in tutte le sue parti, in vigore nel corso di due anni, termine previsto per un efficace e completo processo di implementazione della nuova disciplina europea per le parti coinvolte.

Il legislatore nazionale, dal canto suo, mosso dalla necessità di un adeguamento veloce della disciplina interna, all'emergere di criticità in merito all'impiego delle tecnologie di *A.I.* non è rimasto indifferente alle vicissitudini europee ed internazionali [35], presentando quello che a tutti gli effetti può essere definito un primo intervento di tipo organico sul tema in ambito nazionale. Un primo disegno di legge sull'impiego dell'*A.I.* è stato presentato nel Consiglio dei ministri del 23 aprile 2024 approvando, su proposta del Presidente Giorgia Meloni e del Ministro Carlo Nordio, un disegno per l'introduzione di disposizioni e la delega al Governo in materia di intelligenza artificiale.

Il disegno di legge, come si può evincere dal comunicato stampa del Governo, «individua criteri regolatori capaci di riequilibrare il rapporto tra le opportunità che offrono le nuove tecnologie e i rischi legati al loro uso improprio, al loro sottoutilizzo o al loro impiego dannoso. Inoltre, introduce norme di principio e disposizioni di settore che, da un lato, promuovano l'utilizzo delle nuove tecnologie per il miglioramento delle condizioni di vita dei cittadini e della coesione sociale e, dall'altro, forniscano soluzioni per la gestione del rischio fondate su una visione antropocentrica.

In quest'ottica, il testo di legge definitivo, 23 settembre 2025 n. 132, entrato in vigore il 10 ottobre, non si sovrappone al Regolamento europeo sull'intelligenza artificiale approvato nel marzo 2024 dal Parlamento Europeo, ma ne accompagna il quadro regolatorio in quegli spazi propri del diritto interno, tenuto conto che il regolamento è impostato su un'architettura di rischi connessi all'uso della intelligenza artificiale (IA)» [36].

---

procedure formali di queste istituzioni. Un trilogo può svolgersi in qualsiasi fase della procedura legislativa con l'obiettivo di risolvere le questioni in sospeso ed è presieduto dal colegislatore che ospita la riunione. Il ruolo della Commissione è quello di mediare tra le parti.

Per approfondimenti si veda il sito ufficiale dell'unione europea.

[35] L'ordine esecutivo firmato dal Presidente Joe Biden il 30 ottobre 2023 delinea la posizione politica e strategica degli Stati Uniti sulla creazione, diffusione e utilizzo dei modelli di AI. L'approccio si basa sull'innovazione responsabile, con un occhio di riguardo per la sicurezza, la privacy dei cittadini, l'equità sociale e la promozione dell'innovazione e della competizione

[36] Per approfondimenti v. il comunicato stampa n. 78 del Consiglio dei ministri del 23 aprile 2024.

Dunque, in sostanza, il disegno e la legge definitiva entrato in vigore, individua cinque ambiti prioritari di intervento: *strategia nazionale, autorità nazionali*, *azioni di promozione*, *tutela del diritto di autore* e *sanzioni penali*.

È stata, inoltre, prevista una ulteriore delega al Governo per l'adeguamento della normativa nazionale al regolamento europeo su materie come l'alfabetizzazione dei cittadini sul tema delle *A.I.*, anche mediante appositi percorsi scolastici ed universitari, nonché apposita formazione per ordini professionali ed operatori del settore [37].

Sarà necessario, per quanto attiene alla materia di interesse specifico e le conseguenze dell'intervento normativo nazionale, un ripensamento della disciplina penalistica per l'adeguamento di reati e sanzioni ad un uso illecito delle tecnologie di *A.I.* da parte di attori malevoli.

5. Quali prospettive per il diritto penale nella fase cognitiva

Alla luce dei ragionamenti precedenti risulta del tutto evidente come un uso sostitutivo dell'*A.I.,* mediante decisioni automatizzate, sull'assunto di possibili risvolti positivi in termini di oggettività, non suscettibilità a pressioni esterne, ma, soprattutto, tempistiche più brevi, non sia affatto scevro da possibili conseguenze negative in termini di rispetto delle garanzie dell'ordinamento democratico.

Innanzitutto, è ragionevole considerare il presupposto oggettivo dato da fatto che le varie *A.I.*, attualmente disponibili, dipendano dalla qualità ed integrità dei dati impiegati nel processo elaborativo.

Dunque, sia negli aspetti processuali che nei precedenti giurisprudenziali, elemento essenziale nell'implementazione di tali sistemi dovrebbe essere la costante ed agevole controllabilità di parametri e dati di riferimento, nonché una costante supervisione umana in funzione di filtro e parametro indefettibile di garanzia e responsabilità soggettiva.

L'arrivo sul mercato delle tecnologie di *A.I.* pone tutta una serie di domande alle quali è difficile rispondere senza prima averne testato concretamente l'impiego in circostante della vita fattuale.

In pochissimo tempo, come per altre tecnologie, d'altronde, sono stati compiuti passi da gigante con sviluppi inimmaginabili e opportunità create dai sistemi in grado di porre serissimi quesiti di natura legale ed etica, altrettanto difficili da dirimere, a cui gli addetti ai lavori sono soggetti nelle rispettive attività di implementazione professionale.

[37] V. sempre il comunicato.

In generale, le migliori discussioni nel comparto giustizia sorgono sulla base di questioni etiche e giuridiche di notevole impatto ed i vari quesiti, ormai, quotidianamente sollevati dalla possibile implementazione delle tecnologie di *A.I.* nel sistema penale, godono di un'attenzione privilegiata in funzione dell'impatto diretto sulla vita di ciascun individuo delle scelte compiute nel complesso meccanismo di accertamento dei reati e conseguente giudizio di colpevolezza.

Laddove, in circostante analoghe, si è resa necessaria la valutazione di opportunità per l'impiego di sistemi in grado di ledere la libertà morale dell'individuo da condizionamenti esterni, si è preferito una esclusione aprioristica del supporto tecnico [38].

Storicamente il legislatore, coerentemente con il sistema di equilibri ordinamentali, ha volutamente escluso la possibilità di impiego di simili strumenti, anche a seguito di una scelta consapevole, nel rispetto delle libertà fondamentali, al fine di non subordinare la ricerca della verità storica, derivante dalla esatta ricostruzione dei fatti, alla ricerca della verità giudiziale, unico obiettivo legittimo del processo penale.

La prevalente obiezione potrebbe sorgere da considerazioni di carattere logico dovute al fatto che diverso è l'assoggettamento del soggetto parte del procedimento penale a strumenti tecnici di pressione, dal supporto della moderna tecnologia all'attività cognitiva del giudice nella valutazione di responsabilità dell'imputato.

D'altro canto, come già accennato, esperienze di tipo non cognitivo, ma prognostico, maturate in altri ordinamenti, hanno mostrato criticità significate in ragione della notevole mole di dati cui i sistemi di *A.I.* sono assoggettati per il loro funzionamento con tutte le ripercussioni in termini di qualità della risposta data, quali *bias*, errori statistici e pregiudizi vari, nonché all'impatto che anche un eventuale controllo umano, se non consapevole o poco informato cui meccanismi di funzionamento può avere in termini di settaggio anteriore o acquisizione del risultato *a posteriori* [39].

Nonostante non sia univocamente accettata a livello internazionale una definizione di intelligenza artificiale [40], è pur vero che, fin dalla nascita del termine

[38] L'esempio classico è l'esclusione *a priori* ed in modo oggettivo, anche contro la volontà del soggetto interessato, delle macchine della verità o di mezzi di accertamento dell'attendibilità delle testimonianze. Al riguardo v. App. Brescia, 15 luglio 2020, n. 1683. V. anche G. GENNARI, *Nuove e vecchie scienze forensi alla prova delle corti*, Santarcangelo di Romagna, 2016, pp. 117-123.

[39] Cfr. A. TVERSKY-D. KAHNEMAN, *Judgment under Uncertainty: Heuristics and Biases*, in *Science*, 1974, New Series, 185, 4157 (Sep. 27, 1974), pp. 1124-1131.

[40] L'*A.I. act* la definisce come «qualsiasi modello di implementazione basato su «una macchina» in grado di «dedurre dall'input che riceve», grazie all'implementazione di sofisticate ca-

negli anni Sessanta ad opera del matematico McCarthy, i processi computazionali riconducibili al fenomeno di apprendimento automatizzato hanno subito l'attenzione della comunità di studiosi intenzionati a carpirne potenzialità e pericoli.

Roger Schank, tra i pionieri del metodo computazionale, tentò già negli anni Ottanta di identificarne i caratteri essenziali nella copresenza di cinque elementi distinti: capacità comunicativa, conoscenza di sé, conoscenza del mondo esterno, comportamento teso ad un fine e alto grado di creatività.

Ne deriva un'impossibilità di rappresentare i sistemi di intelligenza artificiale come agenti dalle sembianze umane nelle manifestazioni esteriori e, a maggior ragione, lungi dell'essere direttamente riconducibile nei caratteri intrinseci alle complessità dei meccanismi della mente umana, soprattutto negli aspetti emotivi e relazionali [41].

La considerazione generalizzata dell'importanza di un controllo umano rappresenta un elemento determinante nella regolamentazione dell'*A.I.*, pertanto, è indispensabile stabilire il grado di monitoraggio, coerentemente alla garanzia di adeguato funzionamento del sistema.

Domanda a cui si renderà necessario dare risposte concrete in forma progressiva, in ragione delle evoluzioni della tecnica sempre più esponenziali e di un progressivo adeguamento alle innumerevoli circostante concrete poste al vaglio della cognizione.

L'intervento, recente, del legislatore italiano, potrebbe dare alcune indicazioni di riferimento, intatti, è già possibile intravvedere l'indirizzo generale cui ragionevolmente dovrà proiettarsi chiunque voglia approcciare al tema dell'implementazione dei sistemi di intelligenza artificiale nella giustizia penale.

Tralasciando gli aspetti sanzionatori o legati alla creazione di circostanze aggravanti ad effetto speciale nel caso di uso illecito dell'*A.I.*, è opportuno soffermarsi sulla previsione di intelligenze artificiali nell'amministrazione della giustizia consentita «esclusivamente per finalità strumentali e di supporto, quindi per l'organizzazione e la semplificazione del lavoro giudiziario nonché

pacità adattive dotate di diversi livelli di autonomia, una serie di dati processabili finalizzati a generare svariati «output» (come, ad esempio, «previsioni, raccomandazioni, contenuti, decisioni», ecc.), suscettibili di influenzare «ambienti fisici o virtuali», tenuto conto di un «insieme di obiettivi espliciti o impliciti» concretamente perseguiti, a seconda che siano programmati direttamente da uno sviluppatore umano, oppure definiti mediante il ricorso a tecniche di autoapprendimento algoritmico».

[41] Gli esperti di *A.I.* preferiscono parlare di razionalità, laddove per "razionalità" si intende la «capacità di scegliere la migliore azione da intraprendere per conseguire un determinato obiettivo alla luce di alcuni criteri di ottimizzazione delle risorse a disposizione». Così S. RUSSELL-P. NORVIG, *Artificial intelligence: A Modern Approach*, Boston, III ed., 2009, p. 36 ss.

per la ricerca giurisprudenziale e dottrinale anche finalizzata all'individuazione di orientamenti interpretativi**.** È sempre riservata al magistrato la decisione sull'interpretazione della legge, la valutazione dei fatti e delle prove e sull'adozione di ogni provvedimento inclusa la sentenza» [42].

D'altro canto già in precedenza, il possibile impiego dei sistemi di intelligenza artificiale nelle decisione giudiziarie in ambito penale aveva allarmato il consiglio europeo al punto da spingerlo ad adottare *la Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di giustizia penale e nei relativi ambienti* [43], un documento elaborato dalla Commissione per l'efficacia della giustizia (*CEPEJ*), il 4 dicembre 2018 nella consapevolezza di un incremento dell'impiego dell'intelligenza artificiale nelle "*moderne società*".

Il documento in questione aveva già evidenziato che «anche se non sono specificamente progettati per essere discriminatori, l'uso di algoritmi basati sull'*A.I.* [...] ha mostrato il rischio di favorire la rinascita di teorie deterministiche a scapito delle teorie dell'individualizzazione della pena» [44].

Per evitare di generare fenomeni discriminatori e automatismi, nel processo penale accusatorio, basato principalmente su prove dichiarative, non dovrebbe mai essere permesso a un ipotetico sistema tecnologico di valutazione delle prove di elaborare in via autonoma aspetti critici dovuti a menzogne, reticente od atteggiamenti emotivi, senza il necessario vaglio e supporto esterno dell'agente umano.

Allo stato, sarebbe veramente improbabile, se non del tutto impossibile per le *A.I.,* l'applicazione della regola di giudizio stabilità dall'art. 533 c.p.p. sull'*oltre ogni ragionevole dubbio,* poiché capaci di funzionare, almeno secondo i potenziali tecnologici odierni,, solo con logiche binarie (bianco-nero; vero-falso; sì-no) o probabilistiche in % [45].

Chi sostiene la possibilità di un supporto dei sistemi di intelligenza artificiale al giudizio cognitivo nell'applicazione delle regole del processo penale muove dal presupposto di un costante sviluppo e miglioramento delle tecnologie in

[42] Così al comunicato n. 78 del Governo sul disegno di legge in materia di intelligenza artificiale.

[43] Si veda S. QUATTROCOLO, *Intelligenza artificiale e giustizia: nella cornice della Carta etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *www.lalegislazionepenale.eu*, 18 dicembre 2018, p. 4.

[44] Rispettivamente pp. 16, 41 e 48 del documento.

[45] V. G. CANZIO, *Il dubbio e la legge*, in *Dir. pen. cont.*, 2018, p. 1 ss.; M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei* risk assessment tools *tra Stati Uniti ed Europa*, *ivi*, 2019, p. 1 ss.; A. NATALE, *op. cit.*, p. 1 ss.; nello stesso fascicolo, v. pure i contributi di C. COSTANZI, *La matematica del processo: oltre le colonne d'Ercole della giustizia penale* e di C. CASTELLI-D. PIANA, *Giustizia predittiva. La qualità della giustizia in due tempi*.

questione, tanto da riuscire a rimediare in breve alle innumerevoli criticità derivanti da una tecnologia non matura, al momento, nelle diverse classificazioni, considerata solo di tipo semplice [46], non in grado di sostituire l'agire umano.

Prima di considerare possibile ed opportuno un affidamento al giudizio di valutazione ad una macchina, nel momento cognitivo di valutazione della responsabilità, è necessario concepire margini ben definiti di operatività del sistema, al fine di evitare distorsioni ed anomalie non in linea con il sistema di valori democratici degli ordinamenti occidentali.

la Risoluzione del Parlamento europeo sulla robotica del 2017 enuncia principio di *trasparenza*, secondo cui deve essere sempre possibile ricostruire la logica alla base di ogni decisione in cui sia coinvolta l'intelligenza artificiale, in particolar modo, qualora tale decisione incida sulla vita di una o più persone coinvolte in giudizio.

Anche se non si può ancora veramente parlare di Intelligenza simile a quella umana, carica di sfaccettature emotive e meccanismi intrinsecamente imperscrutabili, è bene tenere alta la guardia nei confronti di fenomeni in costante evoluzione, soggetti ad accelerazioni difficilmente confinabili negli schemi tipici della sovranità democratica.

6. Il giudizio di colpevolezza ed il nesso di causalità (leggi scientifiche di copertura)

Innanzitutto, è necessaria la premessa secondo cui la grande differenza tra i modelli di *A.I.*, sviluppati nella seconda metà del secolo scorso, essenzialmente di tipo simbolico-deduttivo, i quali erano meri esecutori delle regole astratte predeterminate in sede di programmazione [47] e le recenti tecnologie di *machine learning*, prioritariamente di tipo induttivo (c.d. *bottom-up*) e probabilistico, risiede nel fatto che questi ultimi, caratterizzanti la maggior parte delle c.d. *A.I.* oggi diffuse, sfruttano i dati forniti in sede di addestramento, riconoscono i *pattern* statistici sottostanti e ne traggono delle generalizzazioni in funzione della

[46] La distinzione classica è tra *ANI, AGI* ed *ASI*, rispettivamente intelligenze artificiali semplici, generali e superiori.

[47] I modelli simbolico-deduttivi di *A.I.* potevano applicare regole di tipo consequenziale (*if-this-then-that rules*), semplificando e velocizzando le attività umane, ma non erano in grado di gestire situazioni di incertezza. Questi modelli, sebbene utilizzati fino a non molto tempo fa, costituiscono quella che potremmo chiamare l'archeologia dell'intelligenza artificiale e sono oggi comunemente chiamati, con un misto di derisione e nostalgia, "*Good Old-Fashioned AI*" (GOFAI). In tal senso v. W.W.C. ALLEN, *Moral Machines: Teaching Robots Right from Wrong*, Oxford, 2009.

mole di dati analizzati e da cui hanno potuto attingere [48].

Non sono nuove considerazioni circa una tendenza anglofona dei risultati di *output* di tali sistemi, non fosse altro che per la generale impostazione ad analizzare dati nella disponibilità di sistemi prettamente concentrati in contesti che comunemente potremmo considerare di tipo occidentale.

Inutile sottolineare quanto tale limite di accesso all'informazione primaria per le tecnologie interessate dall'elaborazione del dato grezzo possa rappresentate una fonte di errore statistico o, peggio ancora, un pregiudizio assoluto verso certi contesti di riferimento [49].

La maggior parte delle attività umane non possono essere incardinate in regole ben definite, ma sono, essenzialmente, il frutto di processi maturati con esperienze vissute ed imitazione di prossimità.

L'esempio classico è il processo di riconoscimento facciale, quotidianamente messo in pratica dall'agire umano in modo intuitivo, sulla base di precedenti esperienze maturate nel corso di secoli di evoluzione, ma che sarebbe molto difficile da implementare in un algoritmo di riconoscimento tal quale.

Quello che, invece, i sistemi di elaborazioni dei dati riescono a svolgere è, più semplicemente, un processo associativo per analogia, sulla base delle precedenti immagini cui il sistema stesso è stato addestrato. Tuttavia nel momento in cui non ha parametri di riferimento, lo stesso sistema computazionale può semplicemente generalizzare distinguendo un volto da un altro solo sulla base del grado e qualità di addestramento cui è stato sottoposto, mediante un apprendimento c.d. *supervisionato* [50].

Un approccio di tipo "causale" richiederebbe almeno la ricostruzione del processo logico astratto alla base della decisione sul caso concreto, ma, d'altro canto, gli algoritmi attuali non sono in grado di fornire regole generali di classificazione che gli consentano di erogare sempre, in tutti i casi, un *output* ragionato.

Per analogia, appunto, tali sistemi, non saprebbero spiegare quali criteri, tra milioni e milioni di dati inseriti nei database di addestramento siano stati determinanti nell'erogazione dell'*output* finale (c.d. *blackbox*).

I sistemi odierni di *machine learning* consentono solo alle *A.I.* di evidenziare

[48] Per approfondimenti v. J. RUSSEL-P. NORVIG., *op. loc. cit.*

[49] «Qualunque cosa sia *l'A.I.* non è neutrale, e nemmeno noi possiamo esserlo» ha scritto D. MCQUILLAN, *Resisting AI: an anti-fascist approach to artificial intelligence*, Bristol, 2022.

[50] V. G.F. ITALIANO-S. CIVITARESE MATTEUCCI-A. PERRUCCI, *Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni. Una introduzione di contesto*, in A. PAJNO-F. DONATI-A. PERRUCCI (a cura di), *op. cit*., pp. 50-51. V. anche P. DOMINGOS, A Few Useful Things to Know about Machine Learning, in *Communication of ACM*, 55, 10 (October 2012), pp. 78-87.

ed individuare, casomai, *pattern* associativi ricorrenti, difettando nel procedimento di elaborazione analogica, con evidenti carenze di comprensione della semantica, sfociando in errori spesso grossolani e macroscopici, come, per esempio, scambiare un uccello per una pistola o una strada per un muro [51].

Variazioni impercettibili nei dati di *imput* possono, come già detto, essere anche causa di gravi errori di valutazione dei sistemi di intelligenza artificiale, i quali non sono, evidentemente, ancora in grado di erogare modelli astratti di associazione causale delle ricorrenze, ma solo associazioni statistiche, anche se, le tecniche più evolute di *machine learning* sono veramente in grado di generare l'illusione di una superiorità della macchina proprio perché, ragionando in termini statistici matematici, le previsioni algoritmiche sembrano essere più accurate e precise di quelle umane.

Si tratta, banalmente, di una forma allucinatoria-cognitiva, dovuta alla percezione miracolistica di tecnologie emerse d'impatto sullo scenario internazionale, alimentate ad arte dal mercantilismo tipico dell'involuzione al maggior profitto nel minor costo possibile, sia di tempo che di risorse, anche mentali, come se la speranza o il desiderio costante fosse quello di affidare il proprio destino ad un esecutore esterno in funzione di surrogato materiale delle nostre responsabilità [52].

Come noto, il concetto di causalità è strettamente legato al contesto di riferimento, in particolare, al tipo di scienza investita della sua applicazione concreta, intrisa di inferenze valide per il tipo di ragionamento implicato negli interrogativi posti al vaglio dell'operatore.

Il contesto del diritto penale è pervaso da una sua unicità, in ragione della funzione svolta, quale strumento per determinare responsabilità dell'attore sottoposto al vaglio ed eventuali profili puntivi, differenziandosi in tal modo dagli obiettivi della causalità nel contesto civilistico.

Il bisogno di certezza della verità storica aleggia prepotentemente come una spada di Damocle sull'agire penalistico e, nella ricostruzione del processo causale della responsabilità, ci si avvicina quasi ad un immanente bisogno filosofico di ricostruzione della verità umana.

---

[51] Cfr. A. ATHALYE-L. ENGSTROM-A. ILYAS-K. KWOK, *Synthesizing Robust Adversarial Examples*, in *Proceedings of the 35th International Conference on Machine Learning*, Stockholm, PMLR 80, 2018.

[52] In tal proposito v. R. GIACCONI, *Una sorta di magia. Intelligenze artificiali, origami, marionette, serpenti: scoprire l'incanto nei laboratori di robotica*, in *Il Tascabile*, 2 marzo 2023, in cui si riportano le considerazioni di Jamie Paik, direttrice del *Reconfigurable Robotics Lab* (RRL) di Losanna: «ogni volta che accendiamo un robot, gli diamo vita e fa quello che ci aspettavamo, è un momento magico, di gioia, euforia. È come se dicessimo sempre una piccola preghiera: fa' che funzioni questa volta, fa' che funzioni questa volta…».

Da ciò ne deriva come sia del tutto normale pensare alla causalità penale in differente modo in ragione del bisogno analitico, discutendo animosamente di causalità materiale, causalità della colpa o causalità nel concorso.

Della causalità materiale, per l'appunto, il Codice penale ne dà una definizione condizionalistica e temperata attraverso la formula dell'art. 41 secondo cui Il «concorso di cause preesistenti o simultanee o sopravvenute, anche se indipendenti dall'azione od omissione del colpevole, non esclude il rapporto di causalità fra l'azione od omissione e l'evento».

Il codice Rocco, dopo i vari orientamenti ottocenteschi sulla causalità di tipo aritmetico [53], ha fatto suo il modello condizionalistico alleggerendone il dogma con fattori eventuali e preponderanti nella ricostruzione dell'evento e, dunque in grado di interrompere il nesso di causalità, arricchendo la cassetta degli attrezzi del giurista chiamato a rispondere al caso concreto, reso scevro da suggestioni intuitive e malcelati soggettivismi.

Tra gli inconvenienti o *vulnera* intrinseci del condizionalismo non si può fare a meno di notare, ai fini di questo ragionamento, il carattere tendenzialmente eversivo della tipicità penale e, allo stesso tempo, un limite difficilmente superabile nei contesti in cui si renda necessario un giudizio basato sulla valutazione della probabilità statistica della causalità omissiva [54].

I fattori c.d. *interruttivi* della causalità, nel ragionamento sulla causalità umana, assumono carattere dirimente e capacità di interrompere il nesso causale solo qualora si tratti di fenomeno assolutamente eccezionali ed imprevedibili rispetto all'agire umano; pertanto, tale approccio interpretativo ha trovato facile sfogo nella prassi giudiziaria, in quanto non preventivamente incardinata in catalogazione di confine e liberamente determinata nei singoli casi concreti [55].

L'evenienza e la consapevolezza per cui, quasi mai ci si trova dinanzi ai fenomeni anzidetti, ha reso indispensabile l'evoluzione concettuale verso teorie proiettate alla consapevolezza del rischio da riversare nella differente constatazione di attribuzione delle competenze specifiche di gestione del rischio medesimo, senza tralasciare la possibile tracimazione condizionante dell'impatto di azioni ulteriori ed esterne, indipendente dalla possibilità di confinamento [56].

Come noto, molto più difficoltosa diviene la ricostruzione del fatto e della relativa concatenazione causale nella valutazione delle responsabilità in cui si

[53] Si veda R. BLAIOTTA, *Causalità giuridica*, Torino, 2010, p. 85 ss.

[54] Sul punto, G.A. DE FRANCESCO, *Modelli scientifici e cultura dei principi nel rapporto di causalità nel diritto penale*, in *Stud. jur.*, 2002, p. 455 ss.

[55] Al riguardo, v. gli studi in onore di ANTOLISEI scritti da AA.VV., Milano, Giuffrè, 1965.

[56] Si veda R. BLAIOTTA, *L'imputazione oggettiva dell'evento alla luce della teoria del rischio*, in *Cass. pen.*, 2007, p. 365 ss.

rende necessario l'impiego di leggi scientifiche di copertura o massime di esperienza a carattere probabilistico, laddove, anche per chiara forma mentis, potrebbe risultare assai ostico un corretto maneggio delle tematiche in questione da parte dell'agente giudicante, tematiche intrinsecamente a carattere epistemologico e filosofico scientifico [57].

L'eccezionale difficoltà nell'impiego di leggi scientifiche probabilistiche, portatrici di profonda incertezza nella ricostruzione del nesso causale, nasce anche dalla constatazione che nel processo penale la ricostruzione del nesso causale ha necessità di partire da premesse certe ed ottenere risposte sicure, contorniate da considerazioni logiche ed espressioni di un certo rigore.

Per accedere alla necessità di certezze condizionalistiche, pur in presenza delle problematiche sopra accennate, si è reso necessario ricorrere all'applicazione pratica della giurisprudenza, nella figura delle sezioni unite, attraverso la sentenza c.d. *Franzese.*

La stessa ha ribadito la componente normativa della causalità omissiva, non dissimile sul piano logico dal paradigma condizionalistico della causalità commissiva, derivandone così l'inaccettabilità di imputazioni fondate sulla probabilità di incidenza delle condotte mancate.

Nella pronuncia in questione, le considerazioni di carattere epistemologico a proposito di induzione, abduzione, probabilità statistica e probabilità logica muovono dall'assunto per cui il processo penale è retto da criteri probatori di tipo essenzialmente inferenziale-induttivo, muovendo dal fatto storico per come si è presentato, giungendo alla più probabile ricostruzione delle stesse secondo uno schema abduttivo; infatti, le giustificazioni di tale ragionamento, sempre secondo i giudici di legittimità, non sono contenute nelle premesse, va vanno ricercate in elementi ulteriori.

Concentrando la propria attenzione sulle dinamiche del fatto specifico, la conseguenza logica è che la ricostruzione del nesso causale e relativa attribuzione di responsabilità deve essere compiuto sulla base delle evidenze disponibili, dimostrato solo se la condotta dell'agente è condizione necessaria dell'evento in base ai relativi canoni della certezza processuale, «alto grado di credibilità razionale o conferma dell'ipotesi formulata sullo specifico fatto da provare» [58], credibilità che può essere supportata da una elevata probabilità logica e confinante con la certezza.

[57] Sulla ricostruzione della causalità probabilistica si veda O. DI GIOVINE, *Il concetto scientifico e il concetto giuridico di probabilità, il grado di certezza da raggiungere nel giudizio sul fatto*, in AA.VV., *La prova scientifica nel processo penale*, Padova, 2007 p.201 ss. ; ID., *Probabilità statistica e probabilità logica nell'accertamento del nesso causale*, in *Cass. pen*., 2008.

[58] Così Cass., sez. un., 11 luglio 2022, n. 30328.

In tal senso non è ipotizzabile sempre il ricorso a leggi scientifiche di copertura o leggi statistiche con coefficienti molto elevati quali argomenti di spiegazione del ragionamento, ma, in taluni casi, uno su tutti la scienza medica, è possibile il ricorso a leggi scientifiche con coefficienti medio-bassi di probabilità o più comunemente a generalizzazioni esperienziali o rilevazioni epidemiologiche.

In uno specifico passaggio della sentenza si può constatare come la probabilità logica non sia riferibile solo alla legge scientifica impiegata, ma, piuttosto, alle dinamiche inferenziali apportate alla ricostruzione probatoria del nesso causale rispetto alle evidenze ed alle circostanze del caso concreto; ne deriva che non è possibile dedurre in via automatica automaticamente dal coefficiente di probabilità statistica della legge impiegata una conferma dell'ipotesi sul nesso di causalità. La probabilità logica, così intesa, rappresenta solo un fattore di supporto per l'ipotesi iniziale, ma, dal momento che oggetto dell'accertamento è sempre un comportamento umano e non un evento delle scienze naturali, deve essere espresso non attraverso coefficienti numerici, ma con giudizi di tipo qualitativo [59].

Nella prassi, d'altronde, si è accreditata l'idea secondo cui la probabilità logica sia un criterio di tipo quantitativo prossimo al massimo, così come è possibile constatare atteggiamenti volti a corroborare ricostruzioni basate su calcoli probabilistici di riscontri probatori tutto sommato assai deboli, soprattutto nelle ricostruzioni delle causalità omissive.

Nella causalità commissiva, come noto, è possibile fare uso di generalizzazioni probabilistiche, purché dalle stesse si possa inferire a una spiegazione del fatto coerente con gli elementi sulla scena e non inficiata da ipotesi causali alternative, così, il c.d. *giudizio controfattuale*, non pone problemi particolari.

Nella causalità omissiva, invece, l'idea della certezza, è resa assai più aleatoria dai coefficienti probabilistici caratterizzati da generalizzazioni inferenziali, con la necessità di ricostruire la catena causale con modalità differenti, laddove nel giudizio controfattuale occorre inserire l'eventuale condotta diligente omessa, in una rappresentazione ti tipo immaginativo e ideale, tenendo in considerazione la possibile presenza di concause ed interazioni nei fenomeni complessi.

La conoscenza dei casi simili, mediante il ricorso alle già citate leggi scientifiche di copertura e generalizzazioni esperienziali, che, in ragione del loro coefficiente di regolarità causale, ne trasferiscono il relativo grado in incertezza nelle conclusioni [60].

---

[59] Così sempre Cass., sez. un., 11 luglio 2022, n. 30328.

[60] Si veda in tal proposito P. GARBOLINO, *Dall'effetto probabile alla causa probabile*, in *Cass. pen*., 2004, p. 2985 ss.; A. DI MARTINO, *Il nesso causale attivato da condotte omissive tra probabilità, certezza e accertamento*, in *PDP*, 2003, p. 63; L. MASERA, *Il modello causale delle Sezioni Unite e la causalità omissiva*, in *Dir. pen. proc*., 2006, p. 498.

Qualora non vi siano informazioni sufficienti o generalizzazioni logicamente orientate, si rende necessario il ricorso ad un processo inferenziale basato, inizialmente, sulle deduzioni rinvenibili e successivamente nell'induzione al caso concreto [61].

A conti fatti, considerato che le generalizzazioni disponibili in chiave deduttiva subiscono l'applicazione in un secondo momento del ragionamento caratterizzato dall'abduzione operata dal giudicante al caso concreto, l'unico strumento a supporto del ragionamento rimane il ricorso al canone dell'elevata probabilità logica, permanendone, però, i relativi rischi di degenerazione in chiave retorica.

7. Il possibile ruolo dell'*A.I.* nell'accertamento del nesso causale

L'accertamento causale nella scienza penalistica assume un ruolo determinante nell'attribuire la paternità dell'evento e contestuale responsabilità all'agente umano, evidenziando esigenze specifiche di razionalità logica, equa e non influenzata da condizionamenti esterni o intuizioni e suggestioni soggettive.

L'architettura codicistica di tipo condizionalistico comporta che di fronte alla valutazione di un fatto avverso, i vari passaggi del giudizio siano compiuti in una concatenazione logica razionale di ricostruzione dell'evento illecito e relativa imputazione soggettiva dell'agente garante del rischio, con le dovute considerazioni in termini di fattori eccentrici insorti nella catena causale, anche in relazione a comportamenti di attori o fenomeni esterni.

Per quel che qui attiene al ragionamento in merito al possibile impiego degli algoritmi di intelligenza artificiale in funzione di supporto all'accertamento della catena causale, di particolare interesse può risultare, proprio in virtù della sua complessità e particolarità assiologica, il giudizio predittivo della causalità omissiva.

Le particolari dinamiche del caso concreto, in un giudizio di questo tipo, possono determinare la ricostruzione in termini di probabilità logica o elevata credibilità razionale, dell'evitabilità di un evento dannoso al netto dell'inserimento della condotta doverosa richiesta all'agente ed omessa.

La giurisprudenza ha diverse volte sollevato le criticità inerenti all'uso del giudizio di credibilità razionale e il relativo gradiente discrezionale, nella ricostruzione di eventi, che, per quanto introiettati e supportati da ragionamenti inferenziali molto rigorosi, rimangono comunque caratterizzati da una struttura ampiamente probabilistica.

[61] Si veda R. BLAIOTTA, *op. cit.*, p. 405 ss.

Alcuni autori – su tutti Kaheman, Sibony, Sunstein – hanno espresso forti perplessità in merito alla possibilità di impiegare sistemi algoritmici di intelligenza artificiale a supporto del giudizio di probabilità, quali strumenti di analisi ponderate in funzione correttiva dell'errore umano [62].

Certo, c'è da rilevare come gli studi in questione appartengano ad un tempo che, seppur non remoto in termini assoluti, risulta assai antiquato in termini relativi; infatti, da allora, la tecnologia in questione ha subito un progresso di accelerazione ed evoluzione tale per cui non è per niente peregrino rivedere considerazioni e certezze relativo al possibile impiego di *A.I.* nel giudizio di probabilità [63].

Un ruolo determinante, nelle diverse considerazioni sul tema, è senza ombra di dubbio giocato dalla professionalità ed apertura critica dell'agente giudicante, il quale, mediante un esercizio equilibrato ed approfondito della giurisdizione, anche in virtù dell'esperienza maturata, può essere in grado di effettuare una ricostruzione dettagliata del fatto, filtrata da ipotesi, corroborazioni e confutazioni del nesso causale.

D'altro canto, la necessità già maturata dalla giurisprudenza di addurre leggi scientifiche di copertura, accreditate nella ricostruzione del nesso eziologico [64], potrebbe seriamente cozzare con l'impostazione di fondo della maggior parte dei sistemi di *A.I.* in circolazione, viste e considerate le ostiche ricostruzioni dei meccanismi alla base del loro funzionamento [65].

---

[62] In tal senso si veda D. KAHEMAN-O. SIBONY-C.R. SUNSTEIN, *Rumore: un difetto del ragionamento umano*, trad. E. Gallitelli, Milano, 2021.

[63] Sul tema v. C. CASTELLI-D. PIANA, *Giusto processo e intelligenza artificiale*, Santarcangelo di Romagna, 2019; J. NIEVA FENOLI, *Intelligenza artificiale e processo*, Torino, 2019; A. CARLEO (a cura di), *Decisione robotica*, Bologna, 2019.

[64] V. Cass., sez. IV, 17 settembre 2010, n. 43786; Cass. sez. IV, 29 gennaio 2013, n. 16237. V. anche AA.VV., *Prova scientifica, ragionamento probatorio e decisione giudiziale*, Napoli, 2015; D. CANALE, *Norme opache. Il ruolo degli esperti nel ragionamento giuridico*, in *Riv. fil. dir.*, 2015, pp. 93-124; R. BARTOLI, *Diritto penale e prova scientifica*, in *Dir. pen. cont.*, 15/2/2018; F. CAPRIOLI, *La scienza "cattiva maestra". Le insidie della prova scientifica nel processo penale*, in *Cass. pen.*, 2008, pp. 3520-3538; G. CARLIZZI, *La valutazione della prova scientifica*, Milano, 2019, pp. 98-113

[65] In tal senso v. D. CASTELVECCHI, *Can We Open the Black Box of AI*?, in *Nature*, 5 ottobre 2016; M. COECKELBERGH, *Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability*, in *26 Science and Engineering Ethics*, p. 2051 ss.; M. EBER-S. NAVAS, *Algorithms and Law*, *Cambridge University Press, 23 July, 2020*; C. FILICETTI, *Sulla definizione di algoritmo* (nota a Cons. St., sez. III, 25 novembre 2021, n. 7891), in *Giust. ins.*, 8 febbraio 2023; R. GIACCONI, *Una sorta di magia. Intelligenze artificiali, origami, marionette, serpenti: scoprire l'incanto nei laboratori di robotic*a, in *Il Tascabile*, 2 marzo 2023; R. GUIDOTTI-A. MONREALE-S. RUGGIERI-F. TURINI-D. PEDRESCHI-F. GIANNOTTI, *A Survey of Methods for Explaining Black Box Models*, in *ACM CS*, 51, 5, 93, 2018, p. 1 ss.

Dunque, nonostante i tentativi del legislatore europeo e nazionale di agevolare la trasparenza sui meccanismi di funzionamento alla base dei sistemi di intelligenza artificiale, alla luce delle considerazioni sui rischi legati alle c.d. *Blackbox*, nonostante sia in linea astratta possibile indicare in modo preciso quali siano gli elementi di *input* ed *output*, allo stato, non è possibile ricostruire in modo preciso il legate probabilistico tra gli stessi e, a conti fatti, manca così la necessaria legge scientifica di copertura richiesta nella valutazione della concatenazione causale.

Il possibile impiego di scatole nere (EDR), in grado assimilare e memorizzare i possibili passaggi compiuti dall'agente nella ricostruzione della catena causale, non eliminerebbe le criticità legate ai *vulnus* conoscitivi delle ricostruzioni algoritmiche né l'arrivo sulla scena di *A.I. reverse engineering* (c.d. *Explainable Artificial*) sarebbe in grado di dirimere le innumerevoli criticità legate all'impossibilità di ricercare una legge scientifica di copertura in grado di spiegare il nesso eziologico tra *input* ed *output* esplicitandone in modo univoco e non generico il rapporto di causa ed effetto [66].

Inutile evidenziare in tal senso i profili di incompatibilità con il processo di ricostruzione del nesso causale per cui è necessariamente richiesto il riscontro con una legge scientifica di copertura comunemente accettata dalla comunità di riferimento, nonostante con le sezioni unite Franzese sia stato introdotto il metodo temperato ed induttivo della presenza di una legge caratterizzata da un gradiente medio basso di probabilità statistica, purché vi siano contestuali elementi probatori convergenti sull'inesistenza di fattori esterni condizionanti (c.d. *Probabilità logica*) [67].

Come noto, gli errori dei sistemi di *machine learning* possono riguardare i codici sorgente, le alimentazioni o la qualità dei dati impiegati nell'addestramento, difficilmente superabili in tempi brevi.

La regolamentazione europea ha definito standard qualitativi dei sistemi di intelligenza artificiale graduandoli in funzione del rischio connesso al loro impiego.

---

[66] V. in tal senso sempre R. GUIDOTTI-A. MONREALE-S. RUGGIERI-F. TURINI-D. PEDRESCHI-F. GIANNOTTI, *op. loc. cit.*; v. anche A.D. SELBST-S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, in *Fordham LR*, 87, 2018, p. 1110 ss.; T. WISCHMEYER-T. RADEMACHER, *Regulating Artificial Intelligence*, Cham, 2020, p. 75 ss.; per una ricostruzione sistematica dei principali lavori in materia di *Explainable Artificial Intelligence* cfr. F. SOVRANO-S. SAPIENZA-M. PALMIRANI-F. VITALI, *Metrics, Explainability and the European AI Act Proposal*, in *J*, 5, 2022, p. 126 ss.

[67] Sul problema della causalità come "origine" di tutte le altre difficoltà applicative del diritto penale nella materia del danno da prodotto v. C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., *passim*; ID., *Il paradigma della colpa nell'età del rischio*, cit., p. 1670 ss.

Nell'*A.I. Act* si parla per l'appunto di sistemi di rischio inaccettabile, alto, medio, e basso, laddove, proprio per i sistemi considerato ad alto rischio, quali potrebbero essere *A.I.* di supporto alle decisioni giudiziarie, è richiesta l'adozione di un «sistema di gestione dei rischi» (art. 9); l'utilizzo di *dataset* «di qualità» (art. 10); la predisposizione di idonea documentazione tecnica (art.11); l'installazione di *event data recorder* sui dispositivi (art. 12); la previsione di meccanismi che garantiscano il controllo di una persona fisica sul funzionamento del sistema (art. 14, c.d. principio *dell'human-in-the-loop*) [68].

La nuova regolamentazione sembrerebbe introdurre criteri di implementazioni delle *A.I.* probabilmente troppo elastici e conseguenti *vulnus* sui requisiti di determinatezza e precisione delle fattispecie colpose, forse più in linea sull'elaborazione di modelli inerenti alla colpevolezza generica costruita sul riferimento al c.d. *agente modello* [69].

All'art. 9, § 4, lett. a), dell'*AI Act*, infatti, si stabilisce che il sistema di gestione dei rischi debba garantire l'eliminazione o la riduzione dei rischi «per quanto possibile», pertanto, è auspicabile l'implementazione di ulteriori standards qualitativi, così come auspicato anche dal *focus group* congiunto (ISO/IEC JTC 1/SC 42), finalizzato alla pubblicazione di standard per lo sviluppo e la produzione di sistemi di intelligenza artificiale, livelli qualitativi delle tecnologie di *A.I.* per garantire, in tema di discriminazione algoritmica, *risk managment*, affidabilità e qualità delle reti neurali impiegate [70].

8. Conclusioni

È già stato detto come l'*A.I. act* escluda la possibilità di impiego esclusivo, senza supervisione umana, delle tecnologie di intelligenza artificiale nelle deci-

[68] In tal senso v. come i sistemi di *A.I.* considerati ad alto rischio siano individuati dall'art. 6 della proposta attraverso menzione espressa e attraverso rinvio al campo di applicazione di normative europee settoriali. L'art. 6, § 2 rinvia ai settori elencati all'allegato III, che comprendono, da un lato, sistemi che possono mettere in pericolo l'incolumità pubblica o la sicurezza fisica delle persone e, dall'altro, sistemi che, se usati in maniera scorretta, possono causare gravi violazioni dei diritti fondamentali. L'*AI Act* prevede una specifica disciplina in materia di *product safety* soltanto in relazione ai suddetti sistemi «ad alto rischio» (titolo III, artt. 6 ss.).

[69] La Commissione Europea, il 5 dicembre 2022, ha pubblicato la bozza di «richiesta di standardizzazione» nei confronti di CEN e CENELEC. Una volta adottata la proposta, gli standard dovrebbero essere pubblicati entro il 31 gennaio 2025 (dunque prima dell'entrata in vigore dell'*AI Act*, che sarà applicabile dopo 24 mesi dall'approvazione, ai sensi dell'art. 85.

[70] In tal senso v. l'*International Organization for Standardization – International Electrotechnical Commission*, ISO/IEC TR 24028:2020, ISO/IEC TR 24027:2021, ISO/IEC TR 24368:2022.

sioni giudiziali, tuttavia, anche alla luce dei costanti sviluppi della tecnologia in questione, non è da escludere un approccio permissivo alle agevolazioni che potrebbero derivare dall'implementazione della tecnologia medesima nella giustizia cognitiva.

Pertanto, è indispensabile rafforzare la cultura della giurisdizione in una società fortemente condizionata dalla modalità digitale, laddove è sempre richiesto l'uso del giudizio critico e ragionato, che dimostri, nell'eventualità in cui si faccia uso delle tecnologie di intelligenza artificiale nel giudizio, non solo un'attenzione adeguata al fatto concreto posto al vaglio del giudicante, ma, anche, un'approfondita conoscenza dei meccanismi alla base del funzionamento di algoritmi sofisticati di elaborazione dei dati sottoposti al vaglio.

Il legislatore nazionale, d'altro canto, non è rimasto indifferente al costante processo evolutivo della ricerca di settore e la recente legge sull'uso dell'intelligenza artificiale sembrerebbe muovere verso un approccio equilibrato, con le dovute riserve rispetto ad un fenomeno in divenire e dai contorni e potenzialità ancora incerti.

Il potere discrezionale del giudice, nel rispetto dei principi fissati dalla Costituzione e non solo, deve essere esercitato in funzione delle particolarità del caso concreto, adeguandone in modo proporzionato la risposta sanzionatoria.

L'idea di un intervento esterno della macchina a supporto del giudizio di attribuzione della responsabilità mediante algoritmi di elaborazione dei dati dal contorno non ben definito, ancor prima del processo di commisurazione ed applicazione della giusta pena, crea non poche riserve e perplessità nelle coscienze di chi ha ravvisato nelle scienze umane e nella cultura processual-penalistica che li alimenta un baluardo insormontabile di prerogative tese a garanzia dell'agire umano, protetto e non scalfito dalle fascinazione moderniste di prevalenza e prevaricazione della macchina sull'uomo.

È possibile ipotizzare un sostanziale supporto tecnico dei sistemi di intelligenza artificiale nello svolgimento delle attività processuali, soprattutto di tipo routinario e protocollare, quale raccolta di materiale bibliografico, giurisprudenziale o burocratico amministrativo in modo più efficiente ed efficace rispetto alla capacità computazionale di alcune tecnologie di analisi dei dati già a disposizione della macchina giudiziaria, ma, pensare ad un progressivo svuotamento del percorso cognitivo di valutazione delle concatenazioni causali nei giudizi di colpevolezza, quale prerogativa assoluta del giudicante, in ragione degli *output* di *A.I.* più o meno affidabili, sembrerebbe essere se non una vera e propria utopia, quantomeno una banale allucinazione.

AI LAW

INTERNATIONAL REVIEW
OF ARTIFICIAL INTELLIGENCE LAW

EU law seeks to tackle AI opacity, but can it? *

Michael Lognoul

Senior legal researcher at the University of Namur
Guest lecturer at University of Artois and Catholic University of Lyon

Abstract

(EN): The article examines the extent to which European Union law regulates the opacity of artificial intelligence systems – especially those based on machine learning – by imposing transparency obligations. The analysis begins with a reconstruction of the dual origin of the causes of opacity: (i) technical factors (software complexity, high dimensionality, and reliance on statistical correlations that are not necessarily intelligible to human beings) and (ii) legal factors linked to trade secrecy, also safeguarded through the protection of commercial secrets. On these premises, the article considers four obligations recently introduced into the EU legal order – Articles 15 and 17 of the Digital Services Act (DSA), Article 5 of the P2B Regulation, and Article 86 of the AI Act – highlighting a progression from "primary" transparency (the provision of information) to more demanding standards of explanation and justification of individual decisions. While, on the one hand, such obligations can effectively limit voluntary opacity stemming from secrecy, on the other hand certain requirements risk coming into tension with structural limits of machine learning, especially where they call for clear logical links between inputs, system functioning, and outputs that are comprehensible to non-technical recipients. In conclusion, the article proposes strengthening an interdisciplinary approach between legal scholars and computer scientists, and consolidating a shared terminology, in order to align regulatory objectives, technical solutions, and demands for accountability.

(ES): El artículo analiza en qué medida el Derecho de la Unión Europea regula la opacidad de los sistemas de inteligencia artificial, en particular los basados en *machine learning*, mediante la imposición de obligaciones de transparencia. El trabajo parte de la reconstrucción del doble origen de las causas de dicha opacidad: (i) factores técnicos (complejidad del *software*, alta dimensionalidad y dependencia de correlaciones estadísticas no necesariamente inteligibles para el ser humano) y (ii) factores jurídicos vinculados al secreto empresarial, protegido también mediante la tutela de los secretos comerciales. Sobre estas premisas, se examinan cuatro obligaciones

* This publication has received financial support of the Horizon 2020 research and innovation program under grant agreement no. 958339 (DENiM) and of the ARIAC by Digital Wallonia4.AI project under grant agreement no. 2010235. The publication only reflects the views of the author and the funding institutions cannot be held responsible for any use which may be made thereof.

introducidas recientemente en el ordenamiento de la UE – los artículos 15 y 17 del Reglamento de Servicios Digitales (DSA), el artículo 5 del Reglamento P2B y el artículo 86 del Reglamento de IA (AI Act) –, poniendo de relieve una progresión: de la transparencia "primaria" (puesta a disposición de información) a estándares más exigentes de explicación y justificación de decisiones individuales. Si, por un lado, tales obligaciones pueden limitar eficazmente la opacidad voluntaria derivada del secreto, por otro, determinadas prescripciones corren el riesgo de entrar en tensión con límites estructurales del *machine learning*, sobre todo cuando exigen nexos lógicos claros entre *input*, funcionamiento del sistema y *output*, comprensibles para destinatarios no técnicos. En conclusión, se propone reforzar un enfoque interdisciplinar entre juristas e informáticos y consolidar una terminología compartida, con el fin de alinear objetivos normativos, soluciones técnicas y exigencias de *accountability*.

(FR): L'article analyse dans quelle mesure le droit de l'Union européenne encadre l'opacité des systèmes d'intelligence artificielle, en particulier ceux fondés sur le *machine learning*, au moyen de l'imposition d'obligations de transparence. L'étude s'ouvre sur la reconstruction de la double origine des causes de l'opacité: (i) des facteurs techniques (complexité du logiciel, haute dimensionnalité et dépendance à des corrélations statistiques pas nécessairement intelligibles pour l'être humain) et (ii) des facteurs juridiques liés au secret d'affaires, également protégé par la sauvegarde des secrets commerciaux. Sur ces prémisses, l'article examine quatre obligations récemment introduites dans l'ordre juridique de l'UE – les articles 15 et 17 du *Digital Services Act* (DSA), l'article 5 du règlement P2B et l'article 86 de l'*AI Act* – en mettant en évidence une progression: d'une transparence «primaire» (mise à disposition d'informations) vers des standards plus exigeants d'explication et de justification des décisions individuelles. Si, d'une part, ces obligations peuvent limiter efficacement l'opacité volontaire découlant du secret, d'autre part certaines prescriptions risquent d'entrer en tension avec les limites structurelles du *machine learning*, surtout lorsqu'elles requièrent des liens logiques clairs entre *imputs*, le fonctionnement du système et *outputs*, compréhensibles par des destinataires non techniques. En conclusion, l'article propose de renforcer une approche interdisciplinaire entre juristes et informaticiens et de consolider une terminologie partagée, afin d'aligner objectifs normatifs, solutions techniques et exigences d'*accountability*.

(DE): Der Beitrag untersucht, inwieweit das Recht der Europäischen Union die Intransparenz von Systemen künstlicher Intelligenz – insbesondere solcher auf Basis von *Machine Learning* – durch die Auferlegung von Transparenzpflichten reguliert. Ausgangspunkt ist die Rekonstruktion des doppelten Ursprungs der Ursachen von Intransparenz: (i) technische Faktoren (Komplexität der Software, hohe Dimensionalität und Abhängigkeit von statistischen Korrelationen, die für den Menschen nicht notwendigerweise nachvollziehbar sind) und (ii) rechtliche Faktoren, die mit dem Geschäftsgeheimnis zusammenhängen, das auch durch den Schutz von Geschäftsgeheimnissen abgesichert wird. Auf dieser Grundlage werden vier jüngst in die EU-Rechtsordnung eingeführte Pflichten untersucht – die Artikel 15 und 17 des *Digital Services Act* (DSA), Artikel 5 der P2B-Verordnung sowie Artikel 86 des *AI Act* – wobei eine Entwicklung herausgearbeitet wird: von „primärer" Transparenz (Bereitstellung von Informationen) hin zu anspruchsvolleren Standards der Erklärung und Rechtfertigung individueller Entscheidungen. Einerseits können diese Pflichten die freiwillige Intransparenz, die aus Geheimhaltung resultiert, wirksam begrenzen; andererseits geraten bestimmte Vorgaben möglicherweise in Spannung zu strukturellen Grenzen des *Machine Learning*, insbesondere wenn sie klare, für nicht-technische Adressaten verständliche logische Zusammenhänge zwischen *Input*, Funktionsweise des Systems und *Output* verlangen. Abschließend wird vorgeschlagen, einen interdisziplinären Ansatz zwischen Juristinnen und Juristen sowie Informatikerinnen und Informatikern zu stärken und eine gemeinsame Terminologie zu konsoli-

dieren, um normative Zielsetzungen, technische Lösungen und Anforderungen an *Accountability* miteinander in Einklang zu bringen.

(PT): O artigo analisa em que medida o direito da União Europeia disciplina a opacidade dos sistemas de inteligência artificial, em particular os baseados em *machine learning*, por meio da imposição de obrigações de transparência. O trabalho parte da reconstrução da dupla origem das causas da opacidade: (i) fatores técnicos (complexidade do software, alta dimensionalidade e dependência de correlações estatísticas nem sempre inteligíveis para o ser humano) e (ii) fatores jurídicos ligados ao segredo empresarial, protegido também pela tutela dos segredos comerciais. Com base nessas premissas, examinam-se quatro obrigações introduzidas recentemente no ordenamento da UE – os artigos 15 e 17 do *Digital Services Act* (DSA), o artigo 5.º do Regulamento P2B e o artigo 86 do *AI Act* – evidenciando uma progressão: da transparência "primária" (disponibilização de informações) para padrões mais exigentes de explicação e justificação de decisões individuais. Se, por um lado, tais obrigações podem limitar de forma eficaz a opacidade voluntária decorrente do segredo, por outro lado certas prescrições correm o risco de entrar em tensão com limites estruturais do *machine learning*, sobretudo quando exigem nexos lógicos claros entre *input*, funcionamento do sistema e *output*, compreensíveis por destinatários não técnicos. Em conclusão, propõe-se reforçar uma abordagem interdisciplinar entre juristas e informáticos e consolidar uma terminologia partilhada, a fim de alinhar objetivos normativos, soluções técnicas e exigências de *accountability*.

(IT): L'articolo analizza in che misura il diritto dell'Unione europea disciplina l'opacità dei sistemi di intelligenza artificiale, in particolare quelli basati su *machine learning*, attraverso l'imposizione di obblighi di trasparenza. Il lavoro muove dalla ricostruzione della duplice origine delle cause dell'opacità: (i) fattori tecnici (complessità del software, alta dimensionalità e dipendenza da correlazioni statistiche non necessariamente intelligibili per l'essere umano) e (ii) fattori giuridici legati al segreto d'impresa, tutelato anche tramite la protezione dei segreti commerciali. Su tali premesse, vengono presi in esame quattro obblighi recentemente introdotti dall'ordinamento UE – artt. 15 e 17 del *Digital Services Act* (DSA), art. 5 del Regolamento P2B e art. 86 dell'*AI Act* – mettendo in luce una progressione: dalla trasparenza "primaria" (messa a disposizione di informazioni) a *standard* più esigenti di spiegazione e giustificazione di decisioni individuali. Se da un lato tali obblighi possono limitare efficacemente l'opacità volontaria derivante dal segreto, dall'altro talune prescrizioni rischiano di entrare in tensione con limiti strutturali del *machine learning*, soprattutto quando richiedono nessi logici chiari tra *input*, funzionamento del sistema e *output*, comprensibili da destinatari non tecnici. In conclusione, si propone di rafforzare un approccio interdisciplinare tra giuristi e informatici e di consolidare una terminologia condivisa, al fine di allineare obiettivi normativi, soluzioni tecniche e istanze di *accountability*.

Summary: 1. Introduction. – 2. The opacity surrounding AI systems. – 3. Transparency of AI under EU law. – 4. Matching legal requirements and technical limitations of machine learning. – 5. Conclusions.

1. Introduction

The deployment of state-of-the-art artificial intelligence (AI) tools is increasing in all business sectors. The rise of generative AI, like ChatGPT, illustrates this tendency. Advanced AI systems enable the automation of a variety of tasks and are deemed efficient and accurate. Yet, the existence of biases, leading to discriminating outcomes, and of other malfunctions of AI systems are frequently reported [1]. These issues are studied in the scientific literature [2] and are widely recognized by public authorities and political decision-makers [3].

Besides, advanced AI systems are often opaque. This is particularly true for machine learning (ML) systems, which are also referred to as *black boxes* [4].

The opacity of these systems, combined with the existence of potential malfunctions when they are used, generates a series of pitfalls. For instance, actors interacting with AI systems may struggle to accept them or trust the results they generate [5]. People subject to automated AI decision-making may be unable to determine whether the resulting decisions are correct and well-founded or should be challenged [6]. The judicial control of litigations arising from the use of opaque systems may also face the difficulty of determining whether an illegal (or otherwise flawed) output was generated [7].

Against that background, the aim of this paper is to explore and discuss the suitability of several recent transparency requirements imposed by European

[1] See, for instance, the list of errors, mistakes and failures attributed to AI by A. DRAPKIN, *AI Gone Wrong: An Updated List of AI Errors, Mistakes and Failures*, updated on 19 January 2026, available at https://tech.co.

[2] Among others, see A. SELBST-S. BAROCAS, *Big Data's Disparate impact*, in *California LR*, 104, 2016, pp. 671-732; or V. HOFMANN-P.R. KALLURI-D. JURAFSKY-S. KING, *Dialect prejudice predicts AI decisions about people's character, employability, and criminality*, March 2024, available at https://arxiv.org.

[3] At the European level, see European Commission, *White paper on AI – a European approach to excellence and trust*, 19 February 2020, COM(2020)65 final, pp. 12-15.

[4] European Commission, Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee, *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics*, 19 February 2020, COM(2020)64 final, p. 9; J. BURRELL, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, in *Big Data & Society*, 3, 2016, pp. 1-2.

[5] Parliamentary mission report (France), *Donner un sens à l'intelligence artificielle: pour une stratégie nationale et européenne*, 2018, p. 142.

[6] European Commission, *White paper on AI – a European approach to excellence and trust*, 19 February 2020, COM(2020)65 final, p. 14. See also ECJ, 21 June 2022, *Ligue des droits humains c/ Conseil des ministres* (case C-817/19), point 195.

[7] European Commission, *op. loc. ult. cit.*

Union (EU) law. These requirements impose various forms of transparency in AI contexts, as they are precisely designed to tackle the opacity surrounding AI tools and their use.

To achieve this goal, the remainder of the paper is structured as follows. In Section II, the main factors that generate opacity around AI are described. Section III analyzes the four transparency obligations imposed by EU law. These requirements stem from the Regulation on a single market for digital services (DSA) [8], the Regulation on promoting fairness and transparency for business users of online intermediation services (P2B) [9], and the Regulation laying down harmonized rules on artificial intelligence (AIA) [10]. For each obligation, the informational content required is discussed, as well as the *ratio legis* pursued.

Section IV highlights the difficulty of reconciling some of the transparency requirements with the intrinsic characteristics of machine learning systems. To help tackle such issues in the future, Section V concludes with a call for more multidisciplinary approach, based on both law and computer science.

2. The opacity surrounding AI systems

The opacity surrounding AI systems, and particularly those based on ML, stems from a series of factors. They are technical (a), but also legal (b). In this Section, a few elements from both categories are highlighted.

(a) Technical factors of opacity

The technical complexity of state-of-the-art ML systems makes it difficult – if not impossible – to understand them, or the data processing that give rise to the results they output [11].

[8] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), *O.J.*, L.277, 27 October 2022 (hereafter, DSA).

[9] Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services, O*.J.*, L.186, 11 July 2019 (hereafter, P2B).

[10] Regulation 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), *O.J.*, 12 July 2024 (hereafter, AIA).

[11] B. VAASSEN, *AI, opacity, and Personal Autonomy*, in *PhT*, 35, 88, 2022, p. 3.

First, AI systems are, by definition, computer programs [12]. Some form of technical expertise is required to understand the architecture, or the principles behind any computer program. The same is true about reading code. Yet, these technical skills are currently not mastered by a significant proportion of the actors using AI systems [13]. This implies, *per se*, a first layer of opacity.

Second, advanced machine learning systems can deal with numbers of features greater than dozens of billions, both in their learning and operating phases [14]. This is referred to as *high dimensionality* [15]. Considering Miller's Law as a reference point, these orders of magnitude are far beyond human cognitive capacities. In essence, Miller's Law states that a human being can simultaneously consider five to nine parameters, in his short-term memory [16]. A second layer of opacity is thus caused by the fact that human mind cannot apprehend so much information.

Third, opacity arises because machine learning systems learn and operate based on complex, often non-linear statistical correlations (*patterns*), which humans do not intuitively perceive nor understand [17]. The use of statistical relationships by these systems happens regardless of the causal relationships (or other logical links) underlying – and explaining – these statistics. Consequently, being aware of the correlations used by an AI system, to achieve a given task, does not enable human beings to make sense of this information [18].

---

[12] On this point, see AIA, Art. 3(1), which defines an AI system as «[…] a machine-based system […]».

[13] J. BURRELL, *op. cit.*, p. 4; A. SELBST-S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, in *Fordham LR*, 87, 2018, pp. 1093-1094.

[14] Y. BATHAEE, *The artificial intelligence black box and the failure of intent and causation*, in *Harvard JLT*, 31, 2, 2018, p. 904. For instance, GPT3.5, the model used by ChatGPT until July 2024 in its free version, could already deal with 175 billion features, according to BBC Science Focus. See A. HUGHES, *ChatGPT: Everything you need to know about OpenAI's GPT-4 tool*, September 2023, available at https://www.sciencefocus.com.

[15] J. BURRELL, *op. cit.*, p. 5.

[16] See G.A. MILLER, *The magical number seven, plus or minus two: some limits on our capacity for processing information*, in *Psychological Review*, 63, 2, 1956, pp. 81-97.

[17] A. SELBST-S. BAROCAS, *op. cit.*, pp. 1094 and 1096-1097; M. PEREL-N. ELKIN-KOREN, *Black box tinkering: beyond disclosure in algorithmic enforcement*, in *Florida LR*, 69, 2017, p. 189.

[18] J. BURRELL, *op. cit.*, p. 10; M. KRISHNAN, *Against interpretability: a Critical Examination of the Interpretability Problem in Machine Learning*, in *PhT*, 33, 2020, p. 494.

(b) Legal factors of opacity

From a legal perspective, the fact of willingly opacifying elements relating to AI systems enables to benefit from a dedicated regime under intellectual property law: trade secrets protection.

When information (i) is secret, (ii) has commercial value because of its secrecy, and (iii) is subject to reasonable measures to keep it secret, it automatically qualifies as a trade secret [19]. No formal procedures are required to obtain the associated legal protection. Trade secrets are protected against certain types of behaviors, under European law. In particular, the unlawful obtaining, use or disclosure of a trade secret are prohibited, and may give rise to precautionary measures and/or to compensation [20].

Given the broad definition of trade secrets, many elements regarding AI systems can potentially be protected. For example, the very existence of an AI system, its algorithms, architecture, technical characteristics and performance measurements, or the model based on which an ML system operates, could be protected as trade secrets [21]. The source code of the computer program in which an AI system is distributed, as well as the datasets used to train a machine learning system, could potentially also benefit from trade secrets protection [22].

The same goes for the degree of inclusion of an AI system in the performance of business practices. For example, the extent to which an AI system intervenes in a decision-making process could, if this information is kept secret and therefore has commercial value, be considered a business secret in itself.

There are two fundamental differences between technical and legal factors of opacity. On the one hand, only the latter is voluntary, in the sense that it is actively pursued by economic actors, themselves encouraged to do so by the legal framework [23]. On the other hand, legal opacity does not necessarily apply to a single object (i.e., an AI system as a whole), but to a variety of elements relating to an AI system, and to its use. The overlap of technical and legal opacities creates a situation that is all the more complex to grasp.

---

[19] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, *O.J.*, L.157, 15 June 2016, Art. 1.1 and 2.1 (hereafter Directive 2016/943).

[20] Directive (UE) 2016/943, Art. 4, 6 and 10.

[21] C. TSCHIDER, *Legal Opacity: Artificial Intelligence's Sticky Wicket*, in *Iowa LR online*, 126, 2021, pp. 134 and 142.

[22] C. TSCHIDER, *op. loc. ult. cit.*

[23] S. LU, *Algorithmic opacity, private accountability, and corporate social disclosure*, in *Vanderbilt JETL*, 23, 2021, pp. 115-117.

3. Transparency of AI under EU law

In law, attempts to tackle (or limit) the opacity of something often rely on imposing transparency obligations [24]. This type of obligation covers a legal category made up of a wide diversity of norms [25]. Some simply require raw information to be made available. This is transparency in the original meaning of the term: allowing someone to *look through* something, by ensuring access to unaltered information. Others require the communication of information that is more refined, detailed or adapted to its recipient. These range from formal explanations to justifications, which are conceptually close to the legal notion of *motivation*.

Taking into account this disparity among transparency obligations, the Section presents in turn four requirements. They mandate increasingly strong transparency about the AI systems and practices targeted.

(a) Article 15 DSA

Article 15 of the DSA requires providers of intermediary services, such as Internet service providers, hosting companies and online platforms, to supply annual transparency reports.

Under this provision, the reports must be made available to the general public, and be clear and easy to understand. Among the elements to be disclosed, a specific point concerns content moderation activities. On this matter, the reports must contain information about «[...] any use made of automated means for the purpose of content moderation, including a qualitative description, a specification of the precise purposes, indicators of the accuracy and the possible rate of error of the automated means used in fulfilling those purposes, and any safeguards applied» [26].

The aim pursued is twofold: providing information on the technical characteristics of AI systems used to moderate content, and making providers of intermediary services accountable for their practices [27].

---

[24] M. PEREL-N. ELKIN-KOREN, *op. cit.*, p. 186.

[25] V. FORTI, *Réflexions sur la transparence dans les contrats de consommation*, in D. CUSTOS (Ed), *Transparency, a governance principle / La transparence, un principe de gouvernance*, Bruxelles, 2014, p. 296. For an in-depth study on the notion of transparency, see D. HEALD, *Varieties of transparency*, in *Transparency: the Key to Better Governance?*, Oxford, 2006, pp. 25-43.

[26] DSA, Art. 15, paragraph 1, point e).

[27] DSA, Recital 49.

Article 15 of the DSA relates to transparency in the primary sense of the term. It has the effect of making available to the public, the information needed to *see through* the content moderation practices of intermediaries. The opacity that European law seeks to tackle touches upon elements that are generally kept secret on a voluntary basis, and not upon opacity caused by the technical characteristics of content moderation systems.

(b) Article 5 P2B

Article 5 of the P2B regulation imposes the provision of explanations about the main ranking criteria used by certain online platforms, in their relations with their professional users.

More precisely, the platforms targeted are providers of online intermediation services (*e.g.*, Amazon) and online search engines (*e.g.*, Google). The explanations benefit businesses, who depend on the platform to access consumers. The information provided must cover «[...] the main parameters determining ranking [...]» [28] of the goods and services offered. Providers of online intermediation services must also specify the reasons for «[...] the relative importance of those main parameters as opposed to other parameters [...]» [29].

Frequently, AI systems based on machine learning are used by online platforms to decide which main parameters to use, or how much weight to give to the various parameters, to optimize the ranking of goods and services [30].

The objectives of the EU legislator are the following. On the one hand, enabling businesses to effectively compare the ranking practices of various platforms. On the other hand, increasing the ability of these same actors to present their goods or services appropriately, to improve their ranking [31].

Article 5 of P2B imposes a higher level of transparency than Article 15 of the DSA. P2B requires not only the transmission of strictly factual information, but also the provision of reasons, and therefore of logical reasoning. The opacity that EU law seeks to tackle finds part of its roots in trade secrets' practices, particularly regarding the disclosure of main ranking parameters. However, another source of that opacity lies in the technical complexity of AI systems. These systems are able to adapt the weight of the (main) ranking parameters, or even to select these parameters themselves, on the basis of complex statistical

[28] P2B, Art. 5, paragraph 1.

[29] P2B, Art. 5, paragraph 1.

[30] F. PASQUALE, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge (Mass)-London, 2015, pp. 78-80.

[31] P2B, Recital 24.

correlations. The platforms must, then, manage to explain why some particular parameters are selected and used, instead of others, by their AI systems [32].

(c) Article 86 AIA

Article 86 of the AIA is entitled «right to an explanation for individual decision-making». In essence, it grants a new right to all persons affected by decisions made by the deployers of certain high-risk AI systems, where the decisions are *mainly* [33] based on the AI outputs, and insofar as the decisions produce legal (or similarly significant) effects *vis-à-vis* affected persons.

This new right consists in obtaining, upon request, «clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken». Interestingly, this is the first provision of Union law to explicitly confer a right to an explanation (*i.e.*, using the term 'explanation' in its wording), at least in the fight against opacity of AI.

However, there are two major limitations to this Article. On the one hand, among the large number of AI systems that qualify as high-risk systems, only a limited subset enters the scope of Article 86 AIA [34]. On the other hand, the right granted is subsidiary, meaning that it will only apply if no other provision of EU law, that imposes the provision of an explanation, is applicable [35].

Furthermore, the final version of this right is less ambitious than that initially proposed by the European Parliament. The latter also included an explanation of the main parameters used to make the decision, and the data used as input by the AI system [36].

Despite these concerns, it should be highlighted that the text does not simply require the provision of a list of the elements indicated in Article 86. Rather, it imposes an explanation of these elements. The aim is to enable those who are

[32] For an in-depth analysis of Art. 5 P2B, and of its implementation by major platforms, see F. SOVRANO-M. LOGNOUL-A. BACCHELLI, *An Empirical Study on Compliance with Ranking Transparency in the Software Documentation of EU Online Platforms*, *ICSE SEIS'24*, ACM, June 2024, available at https://dl.acm.org.

[33] AIA, Recital 171.

[34] This provision only applies to high-risk systems listed in Annex III (except for its point 2), which excludes high-risk systems listed in Annex II, and Annex III, point 2.

[35] AIA, Art. 86, paragraph 3.

[36] See Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, available at https://www.europarl.europa.eu. More precisely, see amendment 630, which creates a new Art. 68 c in the text.

affected by decisions entering the scope of the Article, to exercise their rights downstream of the explanation provided (for example, to contest the decisions at deployers' level, or to engage into judicial litigations) [37].

As with DSA's Art. 15 and P2B, the opacity that the EU seeks to counter through the AIA stems partially from voluntary business secrecy. This is the case, for instance, regarding the role of AI systems in decision-making processes. However, another part of the opacity fought against by Article 86 AIA originates from the complexity of the AI systems targeted. Explaining the main elements of the decisions taken implies, for deployers, managing to discover the relevant elements considered by their systems, and to relate them to the decisions adopted, despite the potential use of complex statistical correlations by their AI systems.

(d) Article 17 DSA

Article 17 of the DSA requires hosting service providers (and therefore also online platforms) to justify themselves, towards the recipients of their services, when they make certain types of individual decisions that negatively impact them.

More specifically, providers of hosting services must issue a clear and specific *statement of reasons*, when restrictions are imposed on content provided by recipients of their services, that is considered unlawful. The restrictions imposed may, for instance, consist in the demonetization of content, the suspension or deletion of the recipient's account, etc.

The statements of reasons need to be precise and fairly detailed. They must include «[…] information on the use made of automated means in taking the decision […]» [38]. Besides, they must contain «[…] a reference to the legal ground relied on and explanations as to why the information is considered to be illegal content on that ground […]» [39].

The aim of this provision is to enable recipients of hosting services to effectively exercise their right to seek redress [40]. Obviously, to enable service's recipients to determine whether to lodge complaints, and if so, to effectively contest adverse decisions, it is necessary for them to be aware of a series of elements, namely those listed in Article 17 DSA.

To reach its purpose, this provision requires even more detailed, specific and elaborated explanations than discussed above (cfr. (a) to (c)). Article 17 imposes a duty to justify, relating to the reason(s) underlying individual decisions, along-

[37] AIA, Recital 171.

[38] DSA, Art. 17, paragraph 3, point c.

[39] DSA, Art. 17, paragraph 3, point d.

[40] DSA, Recital 54 and Art. 17, paragraph 4.

side with the identification of a legal basis for asserting the unlawfulness of given contents [41]. Such statements of reasons are, ultimately, close to the legal notion of *motivation* [42].

This provision is meant to tackle part of the opacity voluntarily introduced by hosting service providers, to protect their commercial interests. Elements regarding the decision-making processes involved, such as the use of automated systems, must be disclosed. Besides, when decisions are fully automated, the explanations to be provided also intend to tackle the opacity resulting from technical characteristics of AI systems. In particular, Article 17 imposes the establishment and provision of logical links between the elements used as inputs by the AI system, and the legal grounds used to qualify a given content as illegal *in fine*.

4. Matching legal requirements and technical limitations of machine learning

At first sight, the imposition of transparency obligations seems to provide a suitable legal toolkit, to fight against the opacity surrounding AI systems. This is particularly true for the opacity caused by voluntary measures, so as to benefit from trade secrets protection. Lawmakers have the power to limit trade secrets protection, and to impose disclosures of information, where they consider that other interests should prevail upon business secrecy and intellectual property [43]. In that regard, there currently is a trend at EU level towards limiting trade secrets and voluntary opacity, in relation to AI systems. This trend, underlined by the four recent obligations studied, must be encouraged.

Nevertheless, the design and adoption of transparency obligations targeting machine learning systems must also consider the technical limitations of this family of technologies. Imposing transparency obligations that can technically not be met and complied with, could result in a *de facto* ban on the use of these opaque, yet powerful, AI systems [44].

---

[41] A. MICHEL-Y. POULLET, *A partir de l'affaire* Kairos*, quelques réflexions sur les débats autour de la modération des contenus en ligne*, in *R.D.T.I.*, 90, 2023, p. 107.

[42] On the notion of motivation, see J.-L. AUTIN, *La motivation des actes administratifs unilatéraux, entre tradition nationale et évolution des droits européens*, in *RFAP*, 137-138, 2011, pp. 85-99. For an analysis of the notion of motivation in the context of AI decision-making, see A. BIBAL-M. LOGNOUL-A. DE STREEL-B. FRÉNAY, *Legal requirements on explainability in machine learning*, in *AIL*, 29, 2021, pp. 153-155.

[43] Directive (EU) 2016/943, Art. 3, paragraph 2.

[44] A. BIBAL- M. LOGNOUL-A. DE STREEL-B. FRÉNAY, *op. cit.*, p. 164.

Indeed, due to the complex and statistical nature of advanced AI systems, it can be excessively difficult (if not impossible) to provide explanations or justifications that meet the highest legal standards [45]. Problems are particularly encountered when legal texts require the provision of logical links, between the numerous elements considered by an AI system as input, and the result provided by the system as output [46]. Besides, an additional constraint must be taken into account: any explanation provided must be understandable to the human being for whom it is intended, potentially with no background in computer programs, AI or data science [47].

Transparency, in the original meaning of the term, can thus generally be achieved with limited efforts, but obligations requiring a formal explanation or justification are far more difficult to meet. This is due to the inherent limitations of advanced machine learning systems, but that concern does not appear to worry EU lawmakers when designing new obligations. One might highlight, for instance, the wording of recent Article 86 of the AIA to illustrate this statement.

5. Conclusion

The opacity surrounding AI, and its use, is a major concern which may have dire consequences. Opacity is generated by both technical and legal factors, resulting in an overlap of complexity and secrecy.

To overcome opacity issues, several recent pieces of EU law impose transparency obligations. Such obligations range from making raw information available, to providing formal explanations, and justifications. However, there are significant discrepancies between the design of some transparency obligations and the technical limitations of machine learning.

Fortunately, technical research in the field of eXplainable AI (XAI) is currently extremely vivid. New solutions are being developed at an incredible pace by the AI community. Nevertheless, much of this research is being carried out without any consideration for transparency obligations, nor for their *ratio legis*. As a result, it is not clear whether certain technical solutions can enhance compliance with transparency requirements, and if so how, despite a clear convergence of interests between the legal and technical fields in this area [48].

[45] J. BURRELL, *op. cit.*, p. 10.

[46] A. SELBST-S. BAROCAS, *op. cit.*, p. 1097; M. KRISHNAN, *op. cit.*, p. 494.

[47] High level expert group on AI, *Ethics Guidelines for Trustworthy Artificial Intelligence*, 2019, p. 22.

[48] F. SOVRANO-M. LOGNOUL-A. BACCHELLI, *op. cit.*, p. 10.

Therefore, this paper calls for a more effective multidisciplinary dialogue. It is urgent to enable lawmakers, legal experts and IT specialists to understand each other, and work together, on two main points.

The first is to define and communicate clearly on the specific objective(s) to be achieved through the imposition of transparency obligations. The second is to jointly determine the best way to be achieve these objectives, by taking into account the constraints inherent in AI when regulating it.

In order to achieve such a multidisciplinary dialogue, the first step will undoubtedly consist in finding a common ground, by establishing a shared terminology, since words such as *interpretable*, *transparency* or *causal explanation* do not necessarily carry the same meaning when used by a lawyer or a computer scientist.

AI LAW

INTERNATIONAL REVIEW
OF ARTIFICIAL INTELLIGENCE LAW

Streamlining EU Digital Regulation While Enforcing Human-Centric Constitutional Safeguards

Nadia Maccabiani

Professoressa associata di Diritto costituzionale e pubblico nell'Università degli Studi di Brescia

Abstract

(EN): This analysis examines selected EU digital regulations and assesses their ability to address the far-reaching legal and constitutional implications of intrusive profiling based on machine learning. Building on the EU's ongoing efforts within the Digital Omnibus Package to streamline existing digital regulations, the study aims to systematise EU provisions which, although currently dispersed across different legal acts, address various technical and legal aspects related to strong inferential processes resulting from machine learning practices. The objective is to find out a regulatory technique that enhances their effectiveness and ensures that they are better suited to providing appropriate constitutional safeguards.

(ES): Este análisis examina determinadas regulaciones digitales de la Unión Europea y evalúa su capacidad para abordar las profundas implicaciones jurídicas y constitucionales de la elaboración de perfiles basada en el aprendizaje automático. Partiendo de los esfuerzos en curso de la UE, en el marco del *Digital Omnibus Package*, por racionalizar la normativa digital vigente, el estudio pretende sistematizar las disposiciones de la UE que, aunque actualmente se encuentran dispersas en distintos actos jurídicos, abordan diversos aspectos técnicos y jurídicos implicados en las técnicas de perfilado mediante aprendizaje automático. El objetivo es reforzar su efectividad y asegurar que estén mejor adaptadas para proporcionar garantías constitucionales adecuadas.

(FR): Cette analyse examine certaines réglementations numériques de l'Union européenne et évalue leur capacité à répondre aux profondes implications juridiques et constitutionnelles du profilage fondé sur l'apprentissage automatique. S'appuyant sur les efforts en cours de l'UE, dans le cadre du *Digital Omnibus Package*, visant à rationaliser la réglementation numérique existante, l'étude entend systématiser les dispositions de l'UE qui, bien qu'actuellement dispersées dans différents actes juridiques, traitent de divers aspects techniques et juridiques liés aux techniques de profilage par apprentissage automatique. L'objectif est d'en renforcer l'efficacité et de veiller à ce qu'elles soient mieux adaptées pour offrir des garanties constitutionnelles appropriées.

(DE): Diese Analyse untersucht ausgewählte digitale Regelungen der Europäischen Union und bewertet deren Fähigkeit, den weitreichenden rechtlichen und verfassungsrechtlichen Implikationen

des auf maschinellem Lernen beruhenden Profilings zu begegnen. Aufbauend auf den laufenden Bemühungen der EU im Rahmen des *Digital Omnibus Package*, die bestehende digitale Regulierung zu straffen, zielt die Studie darauf ab, EU-Bestimmungen zu systematisieren, die – obwohl derzeit auf verschiedene Rechtsakte verteilt – unterschiedliche technische und rechtliche Aspekte betreffen, die mit *Profiling*-Techniken des maschinellen Lernens verbunden sind. Ziel ist es, ihre Wirksamkeit zu erhöhen und sicherzustellen, dass sie besser geeignet sind, angemessene verfassungsrechtliche Schutzvorkehrungen zu gewährleisten.

(PT): Esta análise examina regulamentos digitais selecionados da União Europeia e avalia a sua capacidade de enfrentar as amplas implicações jurídicas e constitucionais da definição de perfis baseada em aprendizagem automática. Com base nos esforços em curso da UE, no âmbito do *Digital Omnibus Package*, para racionalizar a regulamentação digital existente, o estudo visa sistematizar as disposições da UE que, embora atualmente dispersas por diferentes atos jurídicos, abordam diversos aspetos técnicos e jurídicos envolvidos nas técnicas de definição de perfis por aprendizagem automática. O objetivo é aumentar a sua eficácia e assegurar que estejam mais bem adequadas para proporcionar salvaguardas constitucionais apropriadas.

(IT): Lo scritto prende in esame alcuni regolamenti dell'Unione europea, valutandone l'adeguatezza giuridico-costituzionale rispetto ad intrusive tecniche di profilazione basate sul *machine-learning*. Preso atto del tentativo di razionalizzare la normativa sul digitale, intrapreso dall'UE con il *Digital Omnibus Package*, lo scritto sistematizza le esistenti disposizioni europee che, sebbene distribuite tra vari atti normativi, affrontano questioni tecniche e giuridiche dei processi inferenziali presupposti alla profilazione. L'obiettivo è di suggerire una tecnica regolatoria che rafforzi le esistenti tutele, all'insegna di adeguate garanzie costituzionali.

Summary: 1. A Brief "Technical" Introduction. – 2. From Technology to Law: What Does Machine Learning Profiling Entail? – 3. How Existing EU Regulations Deal with the Issue – 4. A Proposal for a "Re-Oriented" Regulatory Approach.

1. A Brief "Technical" Introduction

The starting point of this research is technical in nature. However, its aim is to translate technical evidence into the legal domain, not in order to assess whether the former complies with the latter, but rather to examine whether the legal framework is adequately structured and provides consistent constitutional safeguards against the risks arising from a technology-enabled evolution that cannot be reversed or erased.

The "technical" point of departure recalls some techniques that have become widespread in the last two decades. These techniques are deployed within the socio-economic and political fields: they not only focus on human cognitive

processes and consequent human behaviours but also target them, in order to steer them towards certain goals. Such techniques, which are termed nudging and neuromarketing, draw on evidence and inferences made by studies conducted within the domain of behavioural sciences and neurosciences. It is not our aim to deepen such practices, but rather to recall them for the limited purposes of the reasoning that will follow. Nudging, based on evidence given by behavioural sciences, exploits human cognitive biases in order to guide human behaviour towards certain aims [1]. Neuromarketing techniques are based on neuroscientific achievements that prove that certain stimuli, even subliminal ones, trigger certain targeted reactions in the human being [2].

Against this backdrop, it is worth recalling a further step within the field of cognitive sciences, with specific reference to computer sciences. This step has been enabled by the development of the internet, global online platforms, the business models of such platforms, and the huge amount of digital data consequently put into circulation, as well as the new artificial intelligence systems that followed, which were no longer knowledge-based and deterministic but were reliant on data and statistical correlations (and were run by machine learning algorithms, with all their sub-categories) [3]. Such machine learning systems, fed with the data, both personal and non-personal data, that people leave during their online and digitally connected experiences, have given rise to the possibility of making insightful and in-depth inferences about people's behaviour and preferences [4]. This technique has been refocused onto the more general practice of profiling. Profiling, legally defined in Article 4, point (4) of the GDPR, es-

[1] For the different kinds of nudging techniques, see R.H. THALER-C.R. SUNSTEIN, *Nudge. The Final Edition*, New York, 2021.

[2] For an overview of the tenets of neuromarketing and the neuroscientific techniques deployed by it, see A. JAVOR-M. KOLLER-N. LEE-L. CHAMBERLAIN-G. RANSMAYR, *Neuromarketing and Consumer Neuroscience: Contributions to Neurology*, in *BMC Neur.*, 13/2013, p. 1 ff.; E. HARRELL, *Neuromarketing: What You Need to Know*, in *Harvard BR*, January 23, 2019; L. BOJIĆ-L. TUCAKOVIĆ-N. NIKOLIĆ, *Neuromarketing Unmasked: A Review of Current State in the Field*, in *Ek. Pred.*, 2021, p. 404.

[3] For a description of machine learning systems as a subset of artificial intelligence systems, see S. RUSSELL-P. NORVIG, *Artificial Intelligence – A Modern Approach*, Hoboken, 2021.

[4] M. HILDEBRANDT, *Defining Profiling: A New Type of Knowledge?*, in M. HILDEBRANDT-S. GUTWIRTH (eds.), *Profiling the European Citizen. Cross-Disciplinary Perspectives*, Dordrecht, 2008, p. 17 ff.; A. MANTELERO, *Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection*, in *CLSR*, 32, 2016, p. 239; L.A. BYGRAVE, *Data Protection Law, Approaching its Rationale, Logic and Limits*, The Hague, 2002, p. 306 ff.; O. SESSO SARTI, *Profilazione e trattamento dei dati personali*, in L. CALIFANO-C. COLAPIETRO (eds.), *Innovazione tecnologica e valore della persona. Il diritto alla protezione dei dati personali nel Regolamento UE 2016/679*, Naples, 2017, p. 585 ff.

sentially entails inferences being extracted from data by deriving certain patterns and predictions in respect of people's behaviours, preferences, emotions and beliefs.

A clear definition of this strong inferential model was given in the 2010 Council of Europe Recommendation, which envisaged profiling as being divided into three stages: «The first stage consists of large-scale collection of data on individual behaviour … During the second stage, these data derived from individual observations undergo computer analysis to correlate certain behavioural characteristics. With statistical tools and algorithms, it thus becomes possible to identify connections between certain kinds of behaviour. Human common sense and logic play no part in establishing these correlations. It is purely the computing power and the sophistication of the algorithms that bring to light correlations often invisible to the naked eye or beyond human reason, albeit without explaining them ... In addition, statistical methods are used to determine a probability factor for the correlation made. In the third stage, the defined correlation is applied to an identified or identifiable individual in order, with a certain margin of error, to deduce some of his or her past, present or future characteristics» [5]. This technique, carried out by machine learning algorithms, improves itself in a circular way, since the system can gather the reaction of the profiled person when subjected to certain stimuli and accordingly fine-tune its inferences on the basis of this recorded reaction, putting in place what computer scientists call «real-time data analysis and optimization» and an «intelligent delivery decision» [6].

Tacking stock of this brief description of the use of machine learning to enable profiling, the "quantum leap" that machine learning has brought about in respect of the traditional inferences made by behavioural sciences and neurosciences, as well as with regard to the kind of stimuli classically offered by nudging and neuromarketing, becomes evident. The causes are multiple. First, in the same way as behavioural science and neuroscience, algorithmically-driven automated profiling is also focused on the behaviour of human beings. However, unlike traditional experiments that first collect and process their evidence "in the lab", machine learning profiling collects and processes its data "outside the lab", constantly following humans' digital interactions in disparate sectors, constantly updating the relevant available data, and constantly making

[5] Recommendation CM/Rec(2010)13 adopted by the Committee of Ministers of the Council of Europe, *The protection of individuals with regard to automatic processing of personal data in the context of profiling*, pars. 96-98.

[6] H. Ji-X. Xu-G. Su-J. Wang-Y. Wang, *Utilizing Machine Learning for Precise Audience Targeting in Data Science and Targeted Advertising*, in *AJST*, 9, 2, 2024, p. 216.

correlations between items of data that stem from such different domains and different human activities [7]. In addition, this is done without provoking the "filtering" conduct that people might engage in when they are aware they are being tested in a lab. Second, as a consequence of two different and intersecting factors, machine learning profiling lays the ground for strengthening the effectiveness of nudging and neuromarketing. On the one hand, the stimuli that behavioural science and neuroscience have proved are able to affect human conduct in the bricks-and-mortar world can now be fine-tuned in a personalised way, using the granular knowledge and understanding of individuals gained by in-depth and detailed machine learning profiling. On the other hand, this fine-tuning can be adapted over time in respect of each single person, according to his/her changing habits and preferences or on the basis of the feedback represented by his/her reaction to the stimuli displayed [8]. In this way, machine learning profiling boosts techniques that already exist, such as nudging and neuromarketing, by relocating them within the realm of the digital world, giving rise to what has been named, according to the terminology of computer sciences, targeting or microtargeting [9]. Unlike bricks-and-mortar nudging and neuromarketing, the pervasiveness of microtargeting stems from people's constant «onlife», and the derived possibility of impinging upon the various fields that have been colonised by the «infosphere» [10] (from the social field to the economic and political relations ones). Third, machine learning profiling can not only boost nudging and neuromarketing, but it also represents the basis and the enabler of further disparate – and potentially harmful – practices (social scoring, emotion recognition, biometric categorisation, etc.).

Bearing all this in mind, our legal and constitutional concern will indeed be focused on the scale and scope of this "quantum leap" undertaken by machine learning profiling, as the next paragraph will explain.

[7] T. COHEN, *Regulating Manipulative Artificial Intelligence*, in *scripted*, 20/2023, p. 205: «Outside the lab, millions of people interact daily with complex machine learning systems designed to 'learn' their behavioural patterns and adapt stimuli (such as newsfeeds and ads) to induce choices which align with the systems' objectives».

[8] K. YEUNG, *"Hypernudge": Big Data as a mode of regulation by design*, in *ICS*, 20, 1/2017, p. 118 ff.; S. MILLS, *Finding the "nudge" in hypernudge*, in *Tech. Soc.*, 71, 2022, p. 150 ff.

[9] H. JI- X. XU-G. SU-J. WANG-Y. WANG, *op. cit*., p. 215 ff.

[10] «Onlife» and «the infosphere» are terms conceived to describe the dominant virtual life of today's humans by the philosopher L. FLORIDI, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Milan, 2017, p. 67 ff.

2. From Technology to Law: What Does Machine Learning Profiling Entail?

What has been briefly described until now represents the "technical" scenario. However, as promised, the purpose of this paper is to shift this scenario into the legal field. In doing so, it becomes evident that what is happening within the domain of cognitive sciences, with specific regard to machine learning enabled profiling, can be considered to be at odds with some tenets of constitutional and legal relevance [11].

This part of the cognitive sciences field (as is typical of the whole cognitive sciences domain) puts its focus on personal aspects and personality traits, and thus on something that is traditionally outside the scope of a democratic and constitutional legal system, since it is treated by such a system as a "holy" space that pertains exclusively to the person [12]. Classically speaking, a democratic and constitutional legal system consistently does not deal with the personal and personality sphere of human beings, except to safeguard it against external coercion, or to address the possibility of it expressing itself, or to allow it to be developed on an informed and fair basis. Even a regulatory technique that is intended to uphold the correct construction of a person's internal will, like disclosure and the consequent information notice as well as the underlying principles of transparency, loyalty and fairness, is aimed at rebalancing existing asymmetries within socio-economic relationships. Thus, once again, the personal sphere of an individual (his/her thoughts and preferences) is relevant for the legal system insofar as it touches upon the external dynamics of the market or of a democratic society. This approach positions the law within the realm of social phenomena, and for this reason puts it in charge of dealing with the external consequences of human behaviour [13].

Our legal concern is not so much grounded on the fact that a person and his/her personal aspects and personality traits is made the object of analysis, calculations, measurements and correlations, but rather on the way that this is

[11] For an inquiry about the impact of quantitative computation on the constitutional protection of the person, see E. DI CARPEGNA BRIVIO, *Pari dignità sociale e* reputation scoring. *Per una lettura costituzionale della società digitale*, Torino, 2024.

[12] J.C. BUBLITZ, *The Nascent Right to Psychological Integrity and Mental Self-Determination*, in A. VON ARNAULD-K. VON DER DECKEN-M. SUSI (eds.), *The Cambridge Handbook of New Human Rights. Recognition, Novelty, Rhetoric*, Cambridge, 2020, p. 387; T. ISTACE, *Protecting the mental realm: What does human rights law bring to the table?*, in *Netherlands QHR*, 41, 4/2023, p. 233.

[13] R. COTTERRELL, *The Sociology of Law. An Introduction,* Butterworths, London, Dublin, Edinburgh, 1992, p. 53 ff.

done by machine learning techniques, which is different from the way that traditional inferences are made in other cognitive science domains [14]. On the one hand, machine learning techniques chart an "inside-out" movement in relation to a person's internal interests, preferences, and emotions, making their prediction its core activity [15]. On the other hand, this understanding and these predictions are leveraged by strong inferential calculations run by machine learning algorithms. By means of these inferences, the knowledge and understanding of a person result in data that are quite detached from the data directly provided by the person or directly observed from his/her conduct. This kind of inference can unravel sensitive aspects of a person even when no special categories of personal data were fed into the machine [16], and can result in personal data even when the input data were not personal [17]. What is involved is not only the personal data that a person has directly delivered to the online platform or digital device while being aware of doing so, nor only the person's observed online conduct and consequent online tracks across different services: these represent the mere starting point of a far more complex process run by sophisticated automated algorithms.

In such cases, it is not only the inside-out movement of their preferences, interests, orientations, and beliefs that skips the will of the data subject: it is in fact the linearity of this movement that is disrupted. Specifically, this disruption occurs because of the intervention of the algorithmically-driven inferences which make the data accomplish the "quantum leap" mentioned above, completely detaching them from the source data. The European Commission has

[14] D. CORACI-I. DOUVEN-G. CEVOLANI, *Inferenza e spiegazione nelle neuroscienze cognitive*, in *Sist. Int.*, 1/2024, p. 25 ff.

[15] S. RODOTÀ, *Tecnopolitica. La democrazia e le nuove tecnologie della comunicazione*, Roma-Bari, 1997, p. 134 ff.

[16] L. EDWARDS-M. VEALE, *Slave to the Algorithm? Why a 'Right to An Explanation' is Probably Not the Remedy You Are Looking For*, in *Duke LTR*, 2017, pp. 36-38. According to Article 29 Working Party, WP251rev.01, Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679, p. 15, profiling «can create special category data by inference from data which is not special category data in its own right but becomes so when combined with other data».

[17] N. LÖLFING.-C. BISCHOFF-BRIEL-A. DIECKHOFF, *Data Protection Law*, in M. ARTZT-O. BELITZ-S. HEMBT-N. LÖLFING (eds.), *International Handbook of AI Law – A Guide to Understanding and Resolving the Legal Challenges of Artificial Intelligence*, The Netherlands, 2025, p. 309: «The autonomous learning process of AI systems can also result in the creation of personal data, where the AI algorithms infers information from anonymous data that ends up becoming personal data by inference. The data might not be considered personal data at the time of collection but is processed by an AI system that creates new connections and analyses to the point where it is possible to infer personal data».

consistently pronounced that the «use of … different types of data may have different implications for the accuracy and the fairness» of profiling practices, «in particular where the processing is opaque or relies on data points whose accuracy is more difficult to be verified» [18]. Thus, as a result of the peculiarity of this process, the distinction between personal and non-personal data is blurred, not only because these types of data are inextricably mixed within the data set, but also because, as mentioned above, even apparently insignificant data can result in personal data (and even special categories of personal data) being produced at the end of the inferential process.

In respect of such situations, the legal protection given through the pattern shaped by privacy and personal data protection rights does not seem to offer adequate safeguards, since the linearity of the connection of the person with his/her data has been broken by a strong inferential process. At the end of the inferential path, the resulting evaluation could be very far removed from the input data and, as such, from the scope of manoeuvre and control of the data subject, even if he/she has been correctly informed (pursuant to the GDPR's provisions) about the automated profiling activity, its purpose and its consequences [19].

All this being considered, it is possible to conclude at this point in time that there is "nothing new under the sun". This situation is something that doctrine has already fully described and examined [20].

Therefore, it is the substance of the inferential process triggered by machine learning profiling that should come into focus, by paying attention to what it actually entails and means in legal terms. In this respect, it should be recalled that making inferences about people implies making "evaluations", and thus "judgements", as stated by the Article 29 Working Party in reference to Article 4, point 4, of the GDPR: «the use of the word 'evaluating' suggests that profiling involves some form of assessment or judgement about a person» [21].

---

[18] EC Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act), par. 159.

[19] See Articles 13, par. 2(f); 14, par. 2(g); 15, par. 1(h) and Article 22 of the GDPR.

[20] As observed by B. CUSTERS, *Profiling as inferred data. Amplifier effects and positive feedback loops*, in E. BAYAMLIOĞLU-I. BARALUIC-L. JANSSENS-M. HILDEBRANDT (eds.), *Being Profiled: Cogitas Ergo Sum. 10 Years of Profiling the European Citizen*, Amsterdam, 2018, p. 113: «The key characteristic of inferred data is that it is data inferred from other data and not data directly or indirectly provided by data subjects»; B. CUSTERS-H. VRABEC, *Tell me something new: data subject rights applied to inferred data and profiles*, in *CLSR*, 52, 2024, p. 11: «it is arguable that in applying analytical techniques data loses the direct connection with the data subject and is thus no longer considered to be 'provided by them'. Rather, it concerns data generated by the data controller».

[21] WP251rev.01 Guidelines on Profiling, cit., p. 7.

However, making judgements about people is not neutral from a legal perspective.

It is not our intention to delve into the issue of algorithmically-driven discrimination [22]: this is a slippery slope since the intrinsically discriminatory nature of profiling surely multiplies the grounds of discrimination [23], and it could lead to the conclusion that each profiled person is – as such – differentiated from others, because of the granularity of insights on people achievable by machine learning profiling [24]. Our intention is to focus on the upstream issue of the inferential process *ex se*, and to recall the "regime" usually followed by the legal system when assessments on people are allowed.

Traditionally speaking, when the legal system explicitly allows assessments and inferences to be made with respect to people, it regulates these "judgements" by means of procedural requirements and ensures that the relevant evaluations are carried out with reference to a limited set of significant, objective and evidence-based data.

More specifically, in the legal system, the assessment or "judgement" of people is usually expressly authorised in pre-established cases, and procedural safeguards are provided by the law. This implies that, before the assessment of a person is permitted, the legal system has already struck a balance between conflicting interests, in compliance with the principle of proportionality and reasonableness [25]. For instance, it is permitted to assess people when fraud, an of-

[22] A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLJ*, 1/2019, p. 63 ff. For new algorithmically-determined categories of discrimination, beyond those provided by EU antidiscrimination law, see B.H.M. CUSTERS, *Reconsidering discrimination grounds in the data economy: an EU comparison of national constitutions*, in *CLSR*, 50, 2023, p. 10. For the distinction between disparate treatment and disparate impact, see S. BAROCAS-A.D. SELBST, *Big Data's Disparate Impact*, in *California LR*, 104, 3/2016, pp. 694-711.

[23] As evidenced by WP251rev.01 Guidelines on Profiling, cit., p. 5, automated profiling has far-reaching implications: «The widespread availability of personal data on the internet and from Internet of Things (IoT) devices, and the ability to find correlations and create links, can allow aspects of an individual's personality or behaviour, interests and habits to be determined, analysed and predicted … Profiling can perpetuate existing stereotypes and social segregation. It can also lock a person into a specific category and restrict them to their suggested preferences. This can undermine their freedom to choose, for example, certain products or services such as books, music or newsfeeds. In some cases, profiling can lead to inaccurate predictions. In other cases, it can lead to denial of services and goods and unjustified discrimination».

[24] B. PARENZO, *Profilazione e discriminazione. Dal GDPR alla Proposta di Regolamento sull'IA*, in *Tecn. dir.*, 2023, p. 106.

[25] As explained by the EC Guidelines on prohibited AI practices, cit., par. 147, a distinction is drawn between lawful and unlawful automated assessments of people: «At the same time, the prohibition is not intended to affect lawful practices that evaluate people for specific purposes

fence or an infringement of the law needs to be proved, when professional/educational competence needs to be tested, when clinical trials are carried out, or when a person's health needs to be checked by the healthcare system.

This results in the enshrinement of certain guarantees which are often composed not only of procedural and formal requirements but also of boundaries, requiring the evaluation be carried out on the basis of objective and verifiable data.

In summary, an assessment can be done by humans or, following technological development, by automated means, for legally envisaged legitimate purposes, in respect of a limited set of objective and verifiable aspects and according to legally pre-settled procedures. As a consistent rule, even when the legal system authorises a process to be automated, it ensures that it is anchored to such limited sets of verifiable and objective data as are relevant in the sector concerned. In this way, the type, scale and scope of the inferences that can be drawn are kept under control.

For instance, automated profiling is allowed to be carried out by financial institutions for anti-money laundering purposes, provided that it is based on the cases, information and procedures envisaged by Chapter III of Regulation (EU) 2024/1624 [26]. Similarly, the Consumer Credit Directive (EU) 2023/2225, when creditworthiness is required to be assessed, makes reference to certain criteria and methods whose aim is to identify relevant and accurate data on the basis of which the evaluation must be carried out [27]. Analogously, inferences made by

that are legitimate and in compliance with Union and national law, in particular where those laws specify the types of data relevant for the specific evaluation purposes and ensure that any resulting detrimental or unfavourable treatment of persons is justified and proportionate».

[26] Art. 76, par. 5 of Regulation (EU) 2024/1624 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing.

[27] Recital 53 makes reference to «information provided by the consumer not only during the preparation of the credit agreement in question, but also during a long-standing commercial relationship». The following Recital 54 clarifies that the «assessment of creditworthiness should be proportionate and done in the interest of the consumer, … and should take into consideration all necessary and relevant factors that could influence a consumer's ability to repay the credit … Member States should be able to issue additional guidance on additional criteria and methods to assess a consumer's creditworthiness, for example by setting limits on loan-to-value or loan-to-income ratios». Recital 55 details that «The assessment of creditworthiness should be based on information on the financial and economic situation. Such information should be necessary and proportionate to the nature, duration, value and risks of the credit for the consumer, in line with the data minimisation principle set out in Regulation (EU) 2016/679, and should be relevant, complete and accurate. That information should include at least the income and expenses of the consumer, including giving appropriate consideration to the consumer's current obligations, inter alia the living expenses of the consumer and the consumer's household, as well as the consumer's financial liabilities. That information should not include special categories of personal

automated data processing on the name records of passengers on a flight, pursuant to Directive (EU) 2016/681, are permitted, according to the interpretation given by the ECJ, provided that they are subject to human revision and are reliant on objective criteria that are objectively linked to the serious crimes addressed by this Directive [28]. Accordingly, the AI Act allows the use of AI systems for the prediction or assessment of the risk of a person committing a crime only when the assessment of the involvement of the person in criminal activity is already based on objective and verifiable facts: in such cases an AI system is permitted to support human assessments [29]. Furthermore, the European Union Fundamental Rights Agency (the FRA), in its guide entitled «Preventing unlawful profiling today and in the future», anchors algorithmic profiling deployed in border management and checks by frontline police officers to objective evidence and reasonable grounds of suspicion [30].

It follows from the above that machine learning profiling disrupts the linearity of the guarantees usually set out by the legal system when the assessment of people is expressly allowed, for a dual reason. On the one hand, it is deeply and strongly inferential, and thus it moves a long way from the original, more objective and verifiable, data [31]. On the other hand, it is subjectively oriented: its purpose is to assess personal aspects and personality traits, and thus, once again, it is not closely linked to objective and verifiable data. In brief, it essentially consists of behavioural analysis for predictive purposes, as stated by the European Data Protection Board (the EDPB) [32].

Thus, although the legal system has usually limited, with adequate and effective safeguards, the types of assessment and judgement that are allowed to be carried out on people, which may be automated, the scale and scope of ma-

data referred to in Article 9(1) of Regulation (EU) 2016/679, such as health data including cancer data, nor information obtained from social networks. The European Banking Authority Guidelines of 29 May 2020 on loan origination and monitoring provide guidelines on what categories of data may be used for the processing of personal data for creditworthiness purposes, which include evidence of income or other sources of repayment, and information on financial assets and liabilities or on other financial commitments».

[28] C-817/19, pars. 219-220 and 259, allows automated data processing in respect of the personal data of people suspected of planning, committing or having committed a serious crime, only where there is objective material from which it can be inferred that those data might, in a specific case, make an effective contribution to combatting such activities.

[29] Article 5, par. 1(d) of the AI Act.

[30] FRA, *Preventing unlawful profiling today and in the future: a guide*, 2018, p. 97 ff.

[31] See footnote 19.

[32] See par. 84 of the EDPB Guidelines 3/2025 on the interplay between the DSA and the GDPR.

chine learning profiling when it is deployed in cases other than those explicitly authorised and regulated skip over these guarantees. This is because the GDPR essentially roots its protection against profiling in the binomial safeguard of transparency and consent, making it possible to remove the prohibition on profiling based solely on automated personal data processing. However, as already evidenced by doctrine, the fundamental rights, freedoms and values at stake are too wide to be adequately safeguarded by means of this individual approach [33]. Therefore, a more structured and comprehensive protection is necessary in order to create an adequate defence in the fast-moving technological evolution that has made profiling algorithms so sophisticated, intrusive, and endowed with strong inferential power.

Although these concerns had been raised by scholars some time ago [34], it is only in recent times that the EU has started to struggle with the issue, complementing the individual approach to automated profiling taken by the GDPR by regulating – in a more efficient way – the high risks and systemic effects which can be triggered by this practice. This raised awareness by the EU is iconically summarised in the ECJ's *Meta Platforms* case, which describes the underlying widespread business model: «based on financing through online advertising, which is tailored to the individual users of the social network according, inter alia, to their consumer behaviour, interests, purchasing power and personal situation. Such advertising is made possible in technical terms by the automated production of detailed profiles in respect of the network users and the users of the online services offered at the level of the Meta group. To that end, in addition to the data provided by the users directly when they sign up for the online services concerned, other user- and device-related data are also collected on and off that social network and the online services provided by the Meta group, and linked to their various user accounts. The aggregate view of the data allows detailed conclusions to be drawn about those users' preferences and interests» [35].

[33] F. BOSCO-N. CREEMERS-V. FERRARIS-D. GUAGNIN-B.J. KOOPS, *Profiling Technologies and Fundamental Rights and Values: Regulatory Challenges and Perspectives from European Data Protection Authorities*, in S. GUTWIRTH-R. LEENES-P. DE HERT (eds.), *Reforming European Data Protection Law*, Dordrecht, 2015, p. 3 ff.

[34] The collective risks of automated profiling became well known because of the Cambridge Analytica scandals, but scholars had proclaimed the threat of the legal implications of this surveillance practice well before: see J. VAN DIJCK, *The culture of connectivity: A critical history of social media*, Oxford, 2013, p. 46 ff.; S. ZUBOFF, *Big other: surveillance capitalism and the prospects of an information civilization*, in *JIT*, 30, 1/2015, p. 75 ff.; F. PASQUALE, *The black box society: The secret algorithms that control money and information*, Cambridge, 2015.

[35] C-252/21, par. 27.

However, it is worth recalling that machine learning profiling does not only underpin this clear business model consisting of algorithmically capturing, assessing and predicting the personal aspects and personality traits of people in order to increase their engagement and keep them stuck to their device as long as possible for advertising goals [36]. It also pervades all the main digital services people use, by shaping the displayed content. From recommender systems at large to newsfeeds, from digital assistants and companions to online games architecture, from cultural to political information, everything can be tailored on the basis of algorithmic inferences.

This is the reason why the legal and constitutional relevance of the great power inherent in machine learning profiling covers both the need to protect the single person and a more collective and systemic perspective. Consequently, the individual stance undertaken by the GDPR certainly represents an important starting point that calls for further complementary provisions that can tackle the collective implications of profiling.

Bearing this in mind, our attention moves now to EU legislation in order to find out how it has evolved in respect of the inferential "technical" process described above.

3. How Existing EU Regulations Deal with the Issue

In respect of automated profiling, the path followed by the EU can be briefly described as swinging between the protection of the single person "from" the market and the protection "of" the market. It can also be briefly described as a path of increasing awareness of the high risks underlying machine learning inferences, with such risks being termed "high" or "systemic".

The bedrock is obviously represented by the provisions of the GDPR, in which respect the other EU legislative acts contain a clause of conformity, usually introduced by the «without prejudice» formula. Article 22 GDPR, dealing

[36] As explained by the European Parliament's resolution of 12 December 2023 on addictive design of online services and consumer protection in the EU single market (2023/2043(INI)), par. A, «whereas in today's attention-based economy, certain technology companies use design and system functionalities to take advantage of users' and consumers' vulnerabilities in order to capture their attention and increase the amount of time they spend on digital platforms; whereas many digital services, such as online games, social media, streaming services for films, series or music, online marketplaces or web shops may be designed to keep users on the platform for as long as possible so as to maximise the data collected and the time and money they spend there as well as to maximise activity, engagement, content production, network development and data sharing».

with fully automated profiling, sets up a regulatory technique that is broadly followed by subsequent and more recent EU interventions. In substance, it enshrines a prohibition, upon the occurrence of certain conditions (i.e. a significant impact on the person), and it foresees a possible derogation upon, among other things, the data subject's consent, provided that certain transparency and due process requirements are met.

Thus, not all automated profiling crosses the threshold of the legal prohibition set out in Article 22. Furthermore, the Working Party and the EDPB explain that fully automated profiling is not considered a technique that *ex se* "significantly affects" a person, but that this judgement depends on the subsequent practices it may enable. For instance, as declared by the Article 29 Working Party, profiling overcomes the threshold (of significantly affecting a person) when it gives rise to targeted advertising based on intrusive insights, «the tracking of individuals across different websites, devices and services», or the exploitation of known vulnerabilities [37]. In this respect, the recent guidelines issued by the EDPB on the interplay between the Digital Services Act and the GDPR have upheld an analogous position [38] and have also added that «it cannot be excluded that the presentation of specific content to users of an online platform via a recommender system would be a 'decision' in the meaning of Article 22(1) GDPR, i.e. a decision which significantly affects the data subject» [39]. In

[37] WP251rev.01 Guidelines on Profiling, cit., p. 22.

[38] EDPB Guidelines No. 3/2025, cit., par. 62: «To assess whether an automated decision to present a specific advertisement to an individual produces legal effects or similarly significantly affects him or her, several (non-exhaustive) characteristics of the personal data processing activity (including at the level of each individual advertisement delivery) should be taken into account, including the intrusiveness of the profiling process, the tracking of individuals across different websites, devices and services; the expectations and wishes of the individuals concerned; the way the advert is delivered; or using knowledge of the vulnerabilities of the data subjects targeted».

[39] EDPB Guidelines 3/2025, cit., par. 84. The following par. 85 explains that «This could be the case, in particular, where the recommender system presents recommendations that cause effects that significantly affect data subjects, i.e. the 'decision' to present specific content to an individual may have an impact that is not necessarily legal but rather economic and social. Particular attention should be paid to cases where algorithmic processes could propose content, services and products that significantly affect individuals having a prolonged or permanent impact on them or significantly affect their behaviour or choices, e.g. recommender systems for housing or job offers on an online platform. The DSA underlines the implications for data subjects related to recommender systems stating that «recommender systems can have a significant impact on the ability of recipients to retrieve and interact with information online, including to facilitate the search of relevant information for recipients of the service and contribute to an improved user experience. They also play an important role in the amplification of certain messages, the viral dissemination of information and the stimulation of online behaviour».

the case of personalised advertising, the legal basis for carrying out automated profiling is narrower than that specified in Article 6, par. 1 of the GDPR. The ECJ, in the *Meta Platforms* case, excluded the performance of a contract and legitimate interest as legal bases. Profiling based solely on automated data processing, pursuant to Article 22 of the GDPR, is also relevant for credit scoring practices. According to the ECJ's adjudication, credit scoring entails the «automated establishment of a probability value based on personal data relating to a person and concerning that person's ability to repay a loan in the future» [40]. Consequently, such a practice, following the ECJ's reasoning, encompasses a profiling activity based solely on automated data processing.

It follows that in all the practices mentioned here (targeted advertising, recommender systems, and credit scoring), due process and transparency duties must be fulfilled; furthermore, meaningful information about the logic of the automated system, and the significance of, and envisaged consequences for, the data subject (Articles 13, par. 2 (f); 14, par. 2 (g); 15, par. 2(h) of the GDPR) [41] must be provided; and the possibility of the data subject expressing his/her point of view, to object the decision and to obtain human revision must be available (Art. 22, par. 3).

Furthermore, for the above-mentioned practices, the most suitable legal basis, in the same way as when special categories of personal data are involved [42],

[40] C-634/21, par. 46.

[41] B. CUSTERS-H. VRABEC, *op. cit*., p. 7, observe, in reference to Articles 13-14 of the GDPR, that «This right to meaningful information about the logic of the profiling may not entail information about the actual profiles or the categories in which data subjects are placed, but it may provide clues about what kinds of profiles or categories are established». In reference to Article 15 of the GDPR, the authors underline that «Even if the inferred data is personal data in the sense that the inferences are ascribed to individuals, it may be unclear whether the right of access can be used by data subjects to get access to such information … This provision does not, however, entail a right for data subjects to receive information about the actual profiles or categories, such as the different types of categories and the category in which the data subject is placed. The provision in Article 15.4 GDPR (stating that all this shall not adversely affect the rights and freedoms of others) could considerably restrict data subject access to categories and profiles. On the one hand, providing such information may interfere with the interests of the data controller, as such information may be considered trade secrets that provide a data controller with competitive edge. It may also create excessive business cost for the data controller to start providing details of how the data has been managed on their end, through which models it has gone and what the analysis has looked like. On the other hand, providing such information may interfere with the interests of other data subjects in the database, as such information may also reveal their characteristics».

[42] C-252/21, par. 89, specified that «where a set of data containing both sensitive data and non-sensitive data is subject to such operations and is, in particular, collected *en bloc* without it being possible to separate the data items from each other at the time of collection, the processing

relies upon the giving of information to the data subject and their explicit consent. However, for this consent to be valid, another condition has been added by the ECJ, pursuant to Article 7, par. 4 of the GDPR: «users are to be offered, if necessary for an appropriate fee, an equivalent alternative not accompanied by such data processing operations [i.e. profiling]» [43].

This approach, grounded on disclosure notices and the consequent awareness of the data subject, reflects the individual stance taken by the GDPR and was adopted in a period when the scale and scope of inferences obtained by machine learning algorithms were not so pervasive and intrusive as they are nowadays.

Transparency duties and explicit consent represent the legal basis for targeted advertising and recommender systems in Regulation (EU) 2022/2065 (DSA) [44]. However, this Act recognises the amplitude of the risks implied by advertising or recommender systems based on machine learning profiling, which extend beyond the sphere of the single person [45], becoming systemic [46]. Evidence of this concern can be found in the express references in the DSA to the use and design of the algorithmic systems at the basis of both recommender systems and the selection and display of advertisements, as factors that influence the severity and probability of the systemic risks [47]. With specific regard to personalised advertisements, the DSA states that «When recipients of the service are presented with advertisements based on targeting techniques opti-

of that set of data must be regarded as being prohibited, within the meaning of Article 9(1) of the GDPR, if it contains at least one sensitive data item and none of the derogations in Article 9(2) of that regulation applies». In addition, special categories of personal data are involved when such information is the result of linking data from different services (pars. 66 ff.).

[43] C-252/21, par. 150.

[44] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act, DSA).

[45] Recital 79 of the DSA is crystal clear about the systemic nature of the risks involved, which is also linked to the platform's business model: «Very large online platforms and very large online search engines can be used in a way that strongly influences safety online, the shaping of public opinion and discourse, as well as online trade. The way they design their services is generally optimised to benefit their often advertising-driven business models and can cause societal concerns».

[46] The DSA divides systemic risks into four categories: (1) dissemination of illegal content (Recital 80); (2) actual or foreseeable impact on the exercise of fundamental rights and freedoms (Recital 81); (3) actual or foreseeable negative effects on democratic processes, civic discourse (Recital 82); (4) «actual or foreseeable negative effect on the protection of public health, minors and serious negative consequences to a person's physical and mental well-being, or on gender-based violence» (Recital 83) and electoral processes, as well as public security.

[47] Article 34, par. 1, and par. 2, points (a) and (d), of the DSA.

mised to match their interests and potentially appeal to their vulnerabilities, this can have particularly serious negative effects. In certain cases, manipulative techniques can negatively impact entire groups and amplify societal harms, for example by contributing to disinformation campaigns or by discriminating against certain groups» [48]. With specific reference to recommender systems, it has been observed that the «DSA recognizes their significant role in the amplification and viral dissemination of content as well as the stimulation of online behaviour» [49].

Taking stock of the severity of such risks, the DSA consistently prohibits the display of advertisements based on profiling using special categories of personal data (Art. 26, par. 3): this is an absolute prohibition, without the possibility of derogation, which is different from what is done in the GDPR [50]. However, a doubt remains: does this prohibition also extend to cases in which special categories of personal data are inferred from data which do not belong to this category? This extension is supported not only by the Working Party position about profiling mentioned above [51], but also by Recital 69 of the DSA, since this includes (going further than the case of profiling using special categories of personal data) the case of a resulting categorisation based on those special categories [52]. In addition, in compliance with the requirement to act in the "best interests of the child", the DSA forbids – in a similar absolute way – the display of advertising based on profiling using the personal data of the recipient of the service when (it is known with reasonable certainty) he/she is a minor (Art. 28).

The DSA has further enhanced the position of the recipient of the service: it has expressly introduced the duty of an online platform to offer an alternative choice. In particular, for targeted advertising (Art. 26, par. 1(d)) and recommender systems (Art. 27, par. 1) the online platform is charged with the duty to

[48] Regulation (EU) 2022/2065, Recital 69.

[49] EDPB Guidelines 3/2025, cit., par. 80.

[50] EDPB Guidelines 3/2025, cit., par. 67: «These special rules laid down by the DSA complement the rules laid down in Article 9(2) GDPR and Article 22(4) GDPR when they apply. The presentation of advertisements based on profiling using special categories of personal data by providers of online platforms to recipients of the service is prohibited by the DSA even in situations where the provider of an online platform or another entity would rely on an appropriate legal basis under Article 6(1) GDPR and an appropriate derogation under Article 9(2) GDPR for this processing».

[51] WP251rev.01 Guidelines on Profiling, cit.

[52] Recital 69 of the DSA states that «providers of online platforms should not present advertisements based on profiling …, using special categories of personal data referred to in Article 9(1) of that Regulation, including by using profiling categories based on those special categories».

provide meaningful information about the main parameters used to determine the recipient to whom the advertisement is presented [53] or the order of the information presented, and about the possibility for the recipient of the service to change these parameters. In addition, Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) are required to provide at least one option, for the main parameters of their recommender systems, which is not based on profiling (Article 39) [54].

However, such transparency requirements are to be fulfilled after the processing of personal data has occurred, as underlined by the recent EDPB Guidelines No. 3/2025. This is different from the information notice to be provided to the data subject under Article 13 of the GDPR, which must be given before the data processing is undertaken [55].

[53] Pursuant to Recital 68 of the DSA, such main parameters include «explanations … on the method used for presenting the advertisement, for example whether it is contextual or other type of advertising, and, where applicable, the main profiling criteria used; it should also inform the recipient about any means available for them to change such criteria».

[54] The EDPB Guidelines 3/2025, cit., par. 87, stipulates that «Providers of VLOPs and VLOSEs may only use a recommender system based on profiling after the recipient of the service has chosen this option. In addition, while the non-profiling based option is active, the provider of the online platform cannot lawfully continue to collect and process personal data to profile the user, for the purposes of future recommendations, e.g. to be prepared in case the user chooses the profiling-based option or to provide more relevant recommendations because they would be based on a more detailed profiling. In addition, where a user uses both versions of a recommender systems, for example by switching several times in the same day, i.e. the version based on profiling and the version not based on profiling, then that user should not be profiled during the use of the version not based on profiling».

[55] EDPB Guidelines No. 3/2025, cit., pars. 51-56, «Article 26(1) DSA lays down transparency rules for providers of online platforms specifically regarding advertising. This provision states that information regarding each specific advertisement should be provided to the recipients of the service in real time. Additionally, meaningful information regarding the main parameters used to determine the recipient to whom the advertisement is presented and, where applicable, information about how to change those parameters as specified in Article 26(1)(d) DSA must be directly accessible from the advertisement. It means that the elements of information mentioned in Article 26 DSA should be provided in real time, directly accessible from the advertisement, while information related to transparency obligations as set out in Articles 13 and 14 GDPR could be presented through the means of a privacy policy (e.g., one click away). It also means that information required under Article 26 DSA would be provided after a processing of personal data may have occurred. This is an important difference between Article 26 DSA and the transparency requirements under the GDPR, since the latter provides that in case of personal data collected directly from the data subject, information shall be provided at the time when personal data are obtained, as set out in Article 13(1) and 13(2) GDPR, before the processing takes place … Moreover, if processing for advertising purposes is based on consent (Article 6(1)(a) GDPR), certain information regarding processing (including profiling) must already be provided to the

When advertising falls under the scope of Article 3, point 2) of Regulation (EU) 2024/900 (TTPA) [56], which contains the definition of political advertising, the threats brought about by profiling can be extensive and impinge upon the correct functioning of democratic processes [57]. Consequently, profiling based on special categories of personal data is forbidden (Article 18, par. 1(c)). In this case, Recital 79 of the TTPA clearly specifies that «profiling using special categories of personal data encompasses profiling using special categories of personal data evaluated from personal data which are not themselves special categories of personal data. This could be the case, for instance, if a data controller uses personal data which are not special categories of personal data to categorise data subjects as having certain religious, philosophical or political beliefs, and regardless of whether that categorisation is true. It should not matter how the category is labelled if the processing of personal data reveals a special category

individual before consent is collected. Additional information pursuant to Article 26(1)DSA would later be directly and easily accessible to the recipient from the advertisement».

[56] Regulation (EU) 2024/900 of the European Parliament and of the Council of 13 March 2024 on the transparency and targeting of political advertising (the TTPA), in its Article 3, point 2), states that «'political advertising' means the preparation, placement, promotion, publication, delivery or dissemination, by any means, of a message, normally provided for remuneration or through in-house activities or as part of a political advertising campaign: (a) by, for or on behalf of a political actor, unless it is of a purely private or a purely commercial nature; or (b) which is liable and designed to influence the outcome of an election or referendum, voting behaviour or a legislative or regulatory process, at Union, national, regional or local level».

[57] The amplitude of such risks is declared in Recital 6 of the TTPA, in reference to political microtargeting: «such techniques may present particular threats to legitimate public interests, such as fairness, equal opportunities and transparency in the electoral process and the fundamental rights to freedom of expression, to privacy and the protection of personal data and to equality and non-discrimination, and the right to be informed in an objective, transparent and pluralistic way»; or, in a more extensive way, by Recital 74 of the TTPA: «Personal data collected directly from individuals, or indirectly such as observed or inferred data, when grouping individuals according to their assumed interests or derived through their online activity, behavioural profiling and other analysis techniques, are increasingly used to target political messages to groups or individual voters or individuals, and to amplify their impact. On the basis of the processing of personal data, in particular special categories of personal data under Regulations (EU) 2016/679 and (EU) 2018/1725, different groups of voters or individuals can be segmented and their characteristics or vulnerabilities exploited, for instance by disseminating the advertisements at specific moments and in specific places, designed to take advantage of the instances where they would be sensitive to a certain kind of information or a message. Such processing of personal data has specific and detrimental effects on individuals' fundamental rights and freedoms, such as to be treated fairly and equally, not to be manipulated, to receive objective information, to form their opinion, to make political decisions and exercise their voting rights. Furthermore, it negatively impacts the democratic process as it leads to fragmentation of the public debate about important societal issues, selective outreach and, ultimately, the manipulation of the electorate».

of personal data» [58]. In addition to this limit, already envisaged by the DSA for general advertisements, a further constraint is added. Profiling is reduced in scope, since only personal data directly collected from the data subject with his/her explicit consent can be processed to display targeted political advertising (Art. 18, par. 1(a) of the TTPA). The introduction of limits to the number of categories of data that can be combined for targeted political advertising is also suggested [59]. Moreover, in order to preserve the validity of consent, the data subject must be offered an equivalent alternative which allows him/her to use the online service without receiving political advertising [60]. However, there is a missed opportunity in respect of similar significant boundaries on profiling: the legislation is focused on what can be defined as political advertising and does not extend to recommender systems that manage the kind and priority of content displayed to the recipient of the service, which can have similar political relevance.

In addition to the above-described protections of data subjects and recipients of services "from" the market, Regulation (EU) 2022/1925 (DMA) [61] lays down provisions addressed to the protection "of" (the correct functioning of) the market [62]. The DMA could be deemed to deal with aspects that (directly and indirectly) draw boundaries on profiling to fulfil both of the following aims: (1) the contestability of the market; and (2) the protection of end users. The former (the contestability of the market) is safeguarded by the DMA's attempts to limit intrusive profiling, since making deep consumer profiling the industry standard would represent a barrier against potential entrants or start-ups because they cannot access data to the same extent, depth, and scale as the gatekeepers [63]. The latter (the protection of end users) is a consequence of this and is upheld by Article 5 of the DMA. More specifically, this Article prevents gatekeepers, in the absence of explicit consent given by the end user, from carrying out the

[58] Recital 79 of the TTPA.

[59] According to Recital 78 of the TTPA: «to help prevent manipulative microtargeting, it is essential that providers of political advertising services take specific measures to ensure that the personal data which is collected and processed for the purpose of targeting and ad delivery of political advertising is limited to what is necessary in relation to that purpose, for instance by restricting the availability of options for targeting and ad delivery of political advertising offered to service recipients to those which require only the combination of up to five categories».

[60] Art. 18, par. 4(b), of the TTPA.

[61] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector (DMA).

[62] M. OROFINO, *Il* Digital Market Act: *una regolazione asimmetrica a cavallo tra diritto alla protezione dei dati e diritto antitrust*, in F. PIZZETTI (ed.), *La regolazione europea della società digitale*, Torino, 2024, p. 175 ff.

[63] See Recital 72 of the DMA.

series of personal data processing practices that are usually necessary for in-depth inferences to form the basis of both personalised advertising (Article 5, par. 1(a)) [64] and personalised content or services (Article 5, par. 1(b)(c)(d)) [65]. The «Joint Guidelines on the Interplay between the DMA and the GDPR», recently adopted by the European Commission and the EDPB, underline that the prohibition on the cross-use of personal data from core platform services in other services and vice versa covers the case of different services provided separately, not the case of services that are provided together or in a mutual supportive way, or third-party services (par. 61) [66]. In this last respect, consent is

[64] Under Article 5, par. 1(a), gatekeepers are prevented from processing, for the purpose of providing online advertising services, personal data of end users using the services of third parties that make use of the core platform services of the gatekeeper.

[65] Pursuant to Article 5, par. 1, gatekeepers are prevented from: (point (b)) combining personal data from the relevant core platform service with personal data from any further core platform services or from any other services provided by the gatekeeper or with personal data from third-party services; (point (c)) cross-using personal data from the relevant core platform service in other services provided separately by the gatekeeper, including other core platform services, and vice versa; and (point (d)) signing in end users to other services of the gatekeeper in order to combine personal data.

[66] According to the Joint Guidelines on the Interplay between the DMA and the GDPR, par. 61: «Article 5(2), point (c) DMA prohibits a gatekeeper from cross-using personal data from a relevant CPS [core platform service] in other services that it provides separately, including other CPSs, and vice versa, without the end user's valid consent. In contrast, Article 5(2), point (c) DMA does not require gatekeepers to obtain consent in instances of cross-use of personal data between a CPS and gatekeeper services that are provided together with or in support of a CPS, or for the cross-use of personal data with third party services». The following par. 67 states that «Article 5(2), point (c) DMA does not inhibit cross-use of personal data that is required to offer the essential functionalities of certain services», such as payment, identification services or online advertising, which «can, in principle, also be considered as services provided together with, or in support of, the gatekeeper's relevant CPS on which ads are displayed». In this last regard, the guidelines underline that a gatekeeper's online search engine CPS can, without consent, cross-use the single search query in its online advertising service, since this may be considered strictly necessary in order to display an ad on the online search engine. In that case, the search query can therefore be cross-used without end user consent under Article 5(2) DMA in order to provide an ad result (par. 70), provided that «processing … [that is] limited [to a] set of on-platform personal data, such as geography (as opposed to precise location), language and content, as well as topics of interest as actively provided by the end user, might not require consent. If processing operations do not involve intrusive measures, such as profiling and tracking, and do not go beyond the reasonable expectations of the end users, it may be possible to rely on Article 6(1), point (f) of the GDPR» (par. 75). However, «the material and temporal scope of the personal data that is cross-used should therefore be limited to what is strictly necessary to offer the interconnected functionalities to cross-use personal data without consent. As a consequence, the retention of personal data that has been cross-used by the gatekeeper should be limited to the time required to carry out the relevant functionality» (par. 69).

not a necessary requirement, and compliance with the GDPR could be achieved through other legal bases, such as the necessity to perform a contract, or legitimate interest [67]. However, «processing operations … [which] involve intrusive measures, such as profiling and tracking» [68] remain forbidden in all cases. Accordingly, when personal data gathered on a platform are cross-used in a supporting service and such processing entails deep profiling activities, consent remains the only legal basis [69]: «certain cross-uses of on-platform data in a supporting advertising service of the gatekeeper may require the latter to obtain consent under the GDPR. This may be the case where such cross-use entails the processing of high volumes and a large variety of types of personal data or involves personal data that allows information falling within Article 9(1) GDPR to be revealed about end users» [70].

The DMA, like the DSA for VLOPs, VLOSEs and their recommender systems, or the TTPA for political advertising, obliges gatekeepers to offer a less personalised alternative, stipulating that this shall not be different from or of lower quality than services provided on the basis of profiling [71].

Furthermore, in order to protect the data subject "from" the market, but also for the protection "of" the correct functioning of the market, Regulation (EU) 2023/2854 (Data Act) [72] states that gatekeepers are excluded from the right to data access and data availability (Article 5, par. 3 and Article 6, par. 2(b)). In addition, third parties who receive data from a data holder upon a request from the user of a connected product or related service are prevented from processing such data for profiling (par. 2(b)). This last prohibition narrows the scope of the

[67] Par. 72 of the Joint Guidelines, cit., underlines that «Provided that their respective conditions are effectively complied with, the lawful grounds of Article 6(1), point (b) or (f) GDPR may be appropriate lawful grounds for the cross-use of personal data between a CPS and another gatekeeper service without end user consent, where both services are provided together with or in support of each other».

[68] Par. 75 of the Joint Guidelines, cit.

[69] As specified in par. 77 of the Joint Guidelines: «certain cross-uses of on-platform data in a supporting advertising service of the gatekeeper may require the latter to obtain consent under the GDPR. This may be the case where such cross-use entails the processing of high volumes and a large variety of types of personal data or involves personal data that allows information falling within Article 9(1) GDPR to be revealed about end users, in a manner that goes beyond their reasonable expectations or that may otherwise have a significant impact on their rights and freedoms».

[70] Par. 77 of the Joint Guidelines, cit.

[71] Recitals 36 and 37 of the DMA and Article 13, par. 6 of the DMA.

[72] Regulation (EU) 2023/2854 of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act).

ability to carry out automated profiling envisaged by Article 22 pars. 2(a) and (c) of the GDPR, since there is no possibility of derogation. Except for this case that "cuts" profiling at its source irrespective of the consequent practice it enables, the above-mentioned legal acts (the GDPR, DSA, DMA and TTPA), when dealing with automated profiling, refer their protection to particular use-cases, primarily targeted advertising and recommender systems. In addition, they essentially build up their safeguards on the requirements of transparency and the explicit consent of the data subject.

In turn, the AI Act defines, in an absolute way and without the possibility of derogation, AI systems that perform profiling on a natural person in the domains listed in Annex III of that Act as high-risk. [73] It then prohibits AI-enabled profiling in reference to some of the listed practices, provided that certain conditions are met. This prohibition encompasses social scoring activities [74], predictive policing or justice [75], inference of emotion [76] in the workplace or an education institution [77], and individual biometric categorisation of sensitive pro-

[73] According to Article 6, par. 3 of the AI Act, «an AI system referred to in Annex III shall always be considered to be high-risk where the AI system performs profiling of natural persons».

[74] Article 5, par. 1(c), of the AI Act prohibits «the placing on the market, the putting into service or the use of AI systems for the evaluation or classification of natural persons or groups of persons over a certain period of time based on their social behaviour or known, inferred or predicted personal or personality characteristics, with the social score leading to either or both of the following: (i) detrimental or unfavourable treatment of certain natural persons or groups of persons in social contexts that are unrelated to the contexts in which the data was originally generated or collected; (ii) detrimental or unfavourable treatment of certain natural persons or groups of persons that is unjustified or disproportionate to their social behaviour or its gravity». As explained by the EC Guidelines on prohibited AI practices, cit., par. 154, since the definition makes reference to the concepts of "evaluation", "inferred characteristics" and "predicted characteristics", it also includes profiling activities.

[75] Article 5, par. 1(d), of the AI Act prohibits «the placing on the market, the putting into service for this specific purpose, or the use of an AI system for making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics; this prohibition shall not apply to AI systems used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity».

[76] As specified by the EC Guidelines on prohibited AI practices, cit., pars. 244-246, «The prohibition in Article 5(1)(f) AI Act does not refer to 'emotion recognition systems', but only to 'AI systems to infer emotions of a natural person'... 'Inferring' is done by deducing information generated by analytical and other processes by the system itself. In such a case, the information about the emotion is not solely based on data collected on the natural person, but it is inferred from other data, including machine learning approaches that learn from data how to detect emotions».

[77] Article 5, par. 1(f), of the AI Act prohibits «the placing on the market, the putting into

files [78]. The lowest common denominator of these forbidden practices is essentially strong subjective-oriented profiling. More specifically, they are cases in which a physical, physiological, behavioural or psychological analysis is carried out in order to make inferences about: (1) personal aspects for scoring (in the case of social scoring); (2) feelings and emotions (in emotion recognition); (3) beliefs and personal orientations (in biometric individual categorisation); or (4) the probability of committing future offences or repeating past offences (in predictive policing and predictive justice).

4. A Proposal for a "Re-Oriented" Regulatory Approach

The European Commission's consultations on the so-called Digital Omnibus Package ran from September 16th, 2025 to October 14th, 2025 [79], and the Commission released its consequent draft proposals on November 19th, 2025 [80]. The

service for this specific purpose, or the use of AI systems to infer emotions of a natural person in the areas of workplace and education institutions, except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons». As explained by the EC Guidelines on prohibited AI practices, cit., par. 249, «Recital 18 AI Act clarifies that emotions or intentions do not include 'physical states, such as pain or fatigue, including, for example, systems used in detecting the state of fatigue of professional pilots or drivers for the purpose of preventing accidents.' It further clarifies that emotion recognition systems do not include 'the mere detection of readily apparent expressions, gestures or movements, unless they are used for identifying or inferring emotions', which should be understood to also apply to Article 5(1)(f) AI Act. Those expressions can be basic facial expressions, such as a frown or a smile, or gestures such as the movement of hands, arms or head, or characteristics of a person's voice, such as a raised voice or whispering. However, when these readily apparent expressions or gestures are used for identifying or inferring emotions or intentions, they are covered by the prohibition».

[78] Article 5, par. 1(g) of the AI Act forbids «the placing on the market, the putting into service for this specific purpose, or the use of biometric categorisation systems that categorise individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation; this prohibition does not cover any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data or categorizing of biometric data in the area of law enforcement». As specified by the EC Guidelines on prohibited AI practices, cit., par. 272: «A wide variety of information, including 'sensitive' information, may be extracted, deduced or inferred from biometric information, even without the knowledge of the persons concerned, to categorise those persons».

[79] See https://ec.europa.eu.

[80] Proposal for a Regulation of the European Parliament and of the Council amending Regulations (EU) 2016/679, (EU) 2018/1724, (EU) 2018/1725, (EU) 2023/2854 and Directives 2002/58/EC, (EU) 2022/2555 and (EU) 2022/2557 as regards the simplification of the digital

declared purpose is to streamline existing EU regulations [81] that indeed demonstrate a piecemeal approach.

We agree with this goal of simplifying the current provisions and re-ordering them in a more systematic way. However, our objective is for the multiple implications of the machine learning profiling of people to be tackled in a better way. Our first concern is its scale and scope: people are constantly within a "digital lab" made of pervasive assessments and judgements, grounded on inferences which have not only moved a long away from a limited set of objective and verifiable data but are also devoid of the traditional ethical and legal guarantees that are usually imposed on experiments conducted within a "real lab" or judgements authorised and regulated by the law. Second, this profiling uses a subjective approach, dealing with the assessment of personal aspects and personality traits, and its results may impinge upon the internal sphere of the individual, a space that a democratic legal system usually deems to be "holy" and "outside the scope" of external interference. The third aspect is its cross-cutting amplitude, since its inferences can leverage different practices (from nudging to neuromarketing to dark patterns to discriminatory decisions), as well as being addressed to different contents or services concerning different domains (from the economic field to the social and political fields).

All this considered, we propose a regulatory technique that takes the critical mass of the achievements already evident in the existing EU legislation with the purpose of fine-tuning them to the multiple challenges posed by the machine learning profiling of people.

The regulatory techniques posited by the legislative acts dealt with in the previous paragraphs, as well as belonging to the area of competition law when oriented to the protection of the correct functioning of the market, enter the

legislative framework, and repealing Regulations (EU) 2018/1807, (EU) 2019/1150, (EU) 2022/868, and Directive (EU) 2019/1024 (Digital Omnibus) - COM(2025) 837 final; Proposal for a Regulation of the European Parliament and of the Council amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI), COM(2025) 837 final.

[81] See section B of the Call for Evidence – Digital Omnibus, where it is states the following as one of the objectives: «Reduce compliance costs for businesses across all sectors in relation to the access, use and sharing of data by reducing fragmentation of rules and their application and by clarifying the rules and requirements that apply and by cutting obligations where a less costly alternative exists. Reduce cookie consent fatigue, strengthen users' privacy rights online, with clear and straight-forward information and options for managing cookies; facilitate the use of cookies and other technologies for businesses for increased data availability. The initiative will pursue a stronger alignment with the general rules on the protection of personal data under the EU data protection law, potentially including modernised rules on cookies under that framework when personal data are collected through such technologies».

realm of consumer protection law when transparency, information notices, valid consent and alternative options are required. In addition, they enter the domain of the regulation of product safety, with transparency, technical requirements, technical documentation, data governance, and testing and monitoring duties being imposed for high-risk artificial intelligence systems. Moreover, they result in the conferral of rights, extending those already provided by the GDPR [82]. Last but not least, a precautionary stance has also been implemented by the EU legislator, by means of the prohibition of some AI practices or the prohibition on making use of special categories of personal data.

Taking stock of this, the proposed regulatory technique consists of an incremental but simplified approach in order to create boundaries for machine learning profiling and push this activity back to the "lowest common denominator" that usually features in the cases, expressly foreseen by the legal system, of authorised judgements and profiling of people. All this should be done without lifting the legal safeguards but by fine-tuning them to address the implications of algorithmically-driven inferences and making them more efficient and effective.

For this purpose, a preliminary consideration needs to be made. This stands on the (already mentioned) "quantum leap" and thus on the distinction between, on the one hand, data that are provided and generated by the data subject and, on the other hand, data that are inferred about people's personal aspects and traits. This is not only a technical distinction: it has received legal status. More specifically, the legal system itself has recognised a demarcation point between data provided and generated by the data subject or the users, on the one hand, and inferred or derived data, on the other. This is the stance adopted by the provisions of the GDPR, the DMA and the Data Act when dealing with the right to data portability, real-time access to data and the right to data availability. In particular, inferred data are outside the scope of these rights, since they are the outcome of an additional and complex algorithmic activity that falls under the domain of the data controller [83], the data holder [84] or the gatekeeper [85]. Consistently, it is the legal system itself that has erected a boundary beyond which data

[82] For instance, as mentioned, the Data Act extends the right to data portability (as does Article 6, par. 9 of the DMA) and introduces the right to real-time, continuous access to data and the right to data availability. The DSA extends the right to receive meaningful information on the logic involved in automated data processing (including profiling), referring to the "main parameters" used in carried out profiling and the consequent right to change them (see the clear explanation given by Recital 68).

[83] Article 20 GDPR.

[84] Articles 3, 4, 5 of the Data Act.

[85] Article 6, par. 9, DMA.

exit the domain of control of the data subject or user and enter the domain of the data holder.

Bearing this in mind, the aim of the regulatory technique we propose is to concentrate on this demarcation point, and the consequent division of the spheres of controllability of data between the data subject and the operator (to use a more neutral term than the ones used in the EU legislation mentioned above), which is the entity who decides on the inferential process for the profiling goals. In this regard, it is worth noting that, in our opinion, the traditional regulatory technique, which relies upon transparency requirements and consent, could be deemed consistent and adequate for placing safeguards to protect data subjects when the data are still within their margin of manoeuvre (and thus when we are dealing with data provided or generated by the data subject). Consequently, we hold that a regulatory technique made of prohibitions and derogations fits the purpose of regulating the procedural aspects that lead to the creation of inferred data, with specific regard to profiling, provided that the "demarcation point" mentioned above is borne in mind.

The baseline of our proposed regulatory technique consists of setting out prohibitions which limit profiling in scale and scope and in this way recover the "lowest common denominator" that makes the data more strongly required to be objective and verifiable; this is what is usually done by a democratic legal system when it explicitly authorises and regulates cases of judgements being made on people, including automated personal data processing (see section 2 above). For this goal to be attained, two intertwined factors are to be taken into consideration, since both contribute to the reduction of in-depth profiling: on the one hand, the categories of data that can be fed into machine learning algorithms; and, on the other hand, the kind of processes these algorithms are allowed to carry out.

As for the former, following the pattern laid down by Article 18, par. 1(a) of the TTPA (whose scope is limited to political advertising), profiling should only be allowed with respect to personal data directly collected from the data subject but excluding special categories of personal data, as also enshrined by the DSA and the TTPA, again in reference to advertising [86]. In this last respect, it should in particular be clearly stated in the binding part of the legal text, and not just in the recitals or in soft law as has been done until now, that the output data of algorithmic inferences are also considered one of the special categories of personal data, even when the input data do not belong to a special category.

As for the latter (the processes run by machine learning algorithms), constraints should be imposed on combining and cross-using data across different

[86] Articles 26, par. 3 and 28 of the DSA. Article 18, par. 1(c) of the TTPA.

services or third-party services, as is done by the DMA. [87] These are preliminary prohibitions, a baseline, as said. Unlike in the model followed by the DMA, the DSA and the AI Act, these constraints should be placed on the "operator", irrespective of the kind of practices triggered by the profiling (targeted advertising, personalised recommender systems, dark patterns, social scoring, emotion recognition, etc.), and irrespective of the operator's size.

This bedrock acquired, it makes sense to recover a further regulatory pattern followed by the EU legislator, which relies upon transparency and consent. Such regulatory safeguards are dispersed and disseminated in different articles in various legislative acts (the GDPR, the DSA, the DMA, the TTPA, the AI Act, and the Data Act) and it would be worth re-ordering them, since they serve the purpose of enhancing the margin of manoeuvre of the data subject *vis-à-vis* machine learning profiling. On the one hand, there is the provision to the data subject, end user or recipient of services (as they are variously called across the EU legislation) of meaningful information about automated profiling, the logic involved, its main parameters, the possibility of changing those parameters, and the significance and envisaged consequences of such processing for the person [88]. On the other hand, there are the requirements for valid consent. The conditions for valid consent have been strengthened over time (from the GDPR, to the DSA, the DMA, and the TTPA, passing through the ECJ's *Meta Platforms* case), by adding the offer to the data subject of an alternative but equivalent option devoid of profiling.

To recap what has been argued, the baseline of prohibitions mentioned above is intended to scale down machine learning profiling and shift it to a process that lies closer to the sphere of checkability, knowledge and understanding of the data subject. This step should make the protection stemming from the information notice and consent more significant and adequate, since the resulting inferences could not undertake the "quantum leap" that detaches them completely from the link with the data subject and the sphere of his/her awareness. The outcome is a closer boundary on inferences, and thus a decreased capability for in-depth insights on people by profiling them. Therefore, the streamlining could be considered to be accomplished without lifting legal safeguards: at the end of the process, automated profiling and the inferences involved would be more strictly related to the data subject, more easily monitored by him/her and also more verifiable and checkable. In this way, the rights-oriented approach

[87] Article 5, pars. 1(a), (b), (c), (d) of the DMA.

[88] See Articles 13, par. 2(f); 14, par. 2(g); 15 par. 1(h); and 22 of the GDPR; see also Article 86 of the AI Act; Articles 26, par. 1(d); 27, pars. 1-2; and 38 of the DSA; Article 18 of the TTPA; Article 5, par. 2 of the DMA.

featured in the GDPR would be recovered in its entire significance.

However, this could not be considered to bring us to the end of the path: a further step should be taken. More specifically, the suggested solution needs to be tested against the principle of necessity and proportionality, and we need to show that a legitimate aim is pursued. The legitimate aim of the above-described prohibition is evident, because of both the individual and the high systemic risks involved in machine learning profiling, which leads to evaluations involving people's personal aspects and personality traits, and in this way impinges upon well-known fundamental freedoms and rights. The means identified to tackle the issue, which are aimed at scaling down the power of inferences, are composed of both the reduction of the set of data fed into a machine learning algorithm and limits on how these data can be processed, as such making the process more verifiable. By acting in this way, the legal system recovers the protections usually erected in permitted cases of judgements and assessments on people.

However, in order to make this solution proportionate, a possibility should be given of a margin of manoeuvre of the data controller, the digital service provider or the provider or deployer of an artificial intelligence system (which have briefly been named "operators"), since they may hold overriding interests underpinning the necessity of in-depth profiling. Therefore, the legal system should provide them with the possibility of lifting the prohibitions erected around machine learning profiling but impose on them the burden of proof. As a result of their technological, economic and/or socio-political power, they occupy the most suitable position to bear this burden of proof. Thus, the choice seems to be reasonable and proportionate and to aim to strengthen the accountability principle, already enshrined in the GDPR [89]. In particular, if they believe they hold a legitimate interest in carrying out machine learning profiling on the basis of huge amounts of data (not only data directly provided by the data subject, but also data observed, derived and inferred, cross-used and combined throughout different services and sources and being personal data of different categories), they should first bear the proof that their interests are of an overwhelming nature as against the rights and freedoms of the data subjects. It is only in this case that judgements and evaluations by strong inferential processes based on personal aspects and personality traits would be permitted. In addition, the preventive authorisation should be given on a case-by-case basis by a competent and independent supervisory authority, which would follow quickly-

[89] S. CALZOLAIO-L. FEROLA-V. FIORILLO-E.A. ROSSI-M. TIMIANI, *La responsabilità e la sicurezza del trattamento*, in L. CALIFANO-C. COLAPIETRO (eds.), *Innovazione tecnologica e valore della persona – Il diritto alla protezione dei dati personali nel Regolamento UE 2016/679*, Napoli, 2017, p. 137 ff.

evolving technological developments and fast-changing socio-economic conditions and avoid the rigidity of a legally enshrined list of practices. This means that the solution is different from that adopted by both the GDPR, under which the assessment of the overriding legitimate interest of the data controller is left to the data controller themselves, and the AI Act, under which the evaluation of a high risk in order to mitigate that risk is left to the provider and deployer of the system.

In this way, the challenging path initiated by the AI Act of listing out forbidden use-cases would be reversed and streamlined: it is not for the legislator to struggle to draft and keep updated such a list, with the implied risk of ineffective provisions due to overinclusive or underinclusive lists or interventions made too late or too early [90]. This is instead a task for the "operator", who holds the necessary technical competences, as well as a more comprehensive knowledge and understanding, to cope with the burden of proving the usefulness, for a specific goal, sector or for society at large, of the automated and in-depth inferences it wishes to undertake on people. Obviously, all the principles that make personal data processing lawful must be respected, including the carrying out of a Data Protection Impact Assessment (DPIA, Article 35 of the GDPR) and a Fundamental Rights Impact Assessment (FRIA, Article 27 of the AI Act) [91], as well as all the requirements set out in Chapter III, Section 2 of the AI Act for high-risk systems.

Last but not least, the legislator obviously retains what is already in place: the pre-emptive right to declare that in certain cases automated inferences are permitted, when the dominance of the public and collective interests has already

[90] In this respect, M. EBERS, *Truly Risk-Based Regulation of Artificial Intelligence. How to Implement the EU's AI Act*, in *The European Union Law Working Paper Series*, Stanford-Vienna, No. 101, 19th June 2024, p. 15, observes that «The AI Act provides a broad and rather abstract classification of high-risk systems under Annex III. Instead of providing a risk classification on a case-by-case basis, the Act uses a pre-defined, closed list of typical high-risk applications. Whether an AI system used in a specific sector for specific purposes poses a high risk to health, safety and/or fundamental rights, is not assessed for the *concrete* risk, but is pre-defined for typical cases in Annex III … The choice of such a top-down regulation raises several issues. First, this approach leads to over-regulation where, for instance, an AI system falls into one of the eight categories listed in Annex III, but in reality does not pose a significant risk of harm. Second, the list of typical high-risk AI systems (albeit with broad definitions and open to updating) may not be easy for the European Commission to keep up to date in a timely manner, given how rapidly AI technology is evolving. Moreover, the decision to delegate (to the Commission) the power to amend Annex III by adding, modifying and removing high-risk AI systems (Art. 7 AI Act) raises concerns in terms of power allocation».

[91] As for the different scopes of, and the consequent relationship between, a DPIA and a FRIA, see O. POLLICINO, *DPIA e FRIA: guida completa alla valutazione del rischio nell'IA*, in *agendadigitale.eu*, 10 November 2025; N. LÖLFING.-C. BISCHOFF-BRIEL-A. DIECKHOFF, *Data Protection Law*, cit., pp. 362-365.

been balanced by the regulator itself, provided that adequate procedural safeguards, as well as relevant, objective and verifiable data to feed the machine learning algorithms, are identified.

In conclusion, since regulatory techniques are not only formal and procedural safeguards but are also substantial guarantees of the effectiveness of constitutional principles, rights and freedoms, it is deemed that the proposed regulatory approach is able to streamline existing EU digital regulations, making them more effective without reducing the legal protection.

AI LAW

INTERNATIONAL REVIEW
OF ARTIFICIAL INTELLIGENCE LAW

Verso un ecosistema digitale europeo integrato. Digital Omnibus, Data Act, AI Act e tutela della persona umana nella governance dei dati e dell'Intelligenza Artificiale

Antonio Felice Uricchio

Ordinario di Diritto tributario
nell'Università degli Studi di Bari Aldo Moro
Presidente ANVUR

Claudio Caldarola

Dottorando di ricerca in Artificial Intelligence Health and Life Sciences
presso l'Università Campus Bio-Medico di Roma
Presidente GP4AI

Abstract

(EN): The recent evolution of the European Union's regulatory framework on data, artificial intelligence, cybersecurity, and digital identity reveals a process of profound institutional transformation, one that goes far beyond the technical dimension of innovation to take on the characteristics of a true constitutional repositioning of the European legal order. Digitalization, far from being merely an instrumental phenomenon, becomes the ground on which the Union reaffirms its political and legal identity, defining a governance model capable of combining technological progress, the integrity of institutions, and the substantial protection of the human person. The result is an innovation law that is not peripheral to the fundamental architecture of the legal order, but is situated at its very center, as it is built on the quality of data, the transparency of algorithms, and public responsibility as an indispensable condition for democratic legitimacy.
The set of recent interventions by the European Union constitutes a broad-reaching project, in which technological innovation, regulatory efficiency, and the protection of the individual are not placed in a dialectical or, much less, a conflicting relationship, but are integrated into a unified conception of digital development as an intrinsically value-driven process. Europe thus commits to building a model where technology does not operate as a factor of restriction of freedoms, but as a tool through which to more fully realize dignity, equality, and democratic participation. In this

sense, digitalization is configured as the place where the capacity of the European legal order to govern its future is measured, while preserving the centrality of the individual and the solidity of institutions. This is the distinctive feature of the path taken: a modernity that advances not against the law, but through the law, and that finds in the rule of law, proportionality, and responsibility the means to ensure that technological progress remains, at all times, irreducibly Human.

(ES): La reciente evolución del marco regulatorio de la Unión Europea en materia de datos, inteligencia artificial, ciberseguridad e identidad digital revela un proceso de profunda transformación institucional, que va mucho más allá de la dimensión técnica de la innovación para asumir las características de un auténtico reposicionamiento constitucional del ordenamiento jurídico europeo. La digitalización, lejos de ser un fenómeno meramente instrumental, se convierte en el terreno sobre el cual la Unión reafirma su identidad política y jurídica, definiendo un modelo de gobernanza capaz de conjugar el progreso tecnológico, la integridad de las instituciones y la protección sustancial de la persona humana. El resultado es un derecho de la innovación que no es periférico respecto de la arquitectura fundamental del ordenamiento, sino que se sitúa en su mismo centro, en cuanto se construye sobre la calidad de los datos, la transparencia de los algoritmos y la responsabilidad pública como condición indispensable de la legitimidad democrática.
El conjunto de las recientes intervenciones de la Unión Europea constituye un proyecto de amplio alcance, en el que la innovación tecnológica, la eficiencia regulatoria y la protección del individuo no se colocan en una relación dialéctica ni, menos aún, conflictiva, sino que se integran en una concepción unitaria del desarrollo digital como proceso intrínsecamente orientado por valores. Europa se compromete así a construir un modelo en el que la tecnología no actúe como un factor de restricción de las libertades, sino como un instrumento a través del cual realizar más plenamente la dignidad, la igualdad y la participación democrática. En este sentido, la digitalización se configura como el ámbito en el que se mide la capacidad del ordenamiento jurídico europeo para gobernar su futuro, preservando al mismo tiempo la centralidad del individuo y la solidez de las instituciones. Este es el rasgo distintivo del camino emprendido: una modernidad que avanza no contra el derecho, sino a través del derecho, y que encuentra en el Estado de derecho, la proporcionalidad y la responsabilidad los medios para garantizar que el progreso tecnológico permanezca, en todo momento, irreductiblemente Humano.

(FR): L'évolution récente du cadre réglementaire de l'Union européenne en matière de données, d'intelligence artificielle, de cybersécurité et d'identité numérique révèle un processus de profonde transformation institutionnelle, qui dépasse largement la seule dimension technique de l'innovation pour revêtir les traits d'un véritable repositionnement constitutionnel de l'ordre juridique européen. La numérisation, loin d'être un phénomène purement instrumental, devient le terrain sur lequel l'Union réaffirme son identité politique et juridique, en définissant un modèle de gouvernance capable d'articuler le progrès technologique, l'intégrité des institutions et la protection substantielle de la personne humaine. Il en résulte un droit de l'innovation qui n'est pas périphérique à l'architecture fondamentale de l'ordre juridique, mais s'en situe au cœur même, puisqu'il se construit sur la qualité des données, la transparence des algorithmes et la responsabilité publique comme condition indispensable de la légitimité démocratique.
L'ensemble des interventions récentes de l'Union européenne constitue un projet de grande envergure, dans lequel l'innovation technologique, l'efficacité normative et la protection de l'individu ne sont pas placées dans une relation dialectique, et encore moins conflictuelle, mais intégrées dans une conception unifiée du développement numérique comme processus intrinsèquement guidé par des valeurs. L'Europe s'engage ainsi à bâtir un modèle où la technologie n'opère pas comme un facteur de restriction des libertés, mais comme un instrument permettant de réaliser plus pleinement la dignité, l'égalité et la participation démocratique. En ce sens, la numérisation se configure comme l'espace où se mesure la capacité de l'ordre juridique européen à gouverner

son avenir, tout en préservant la centralité de l'individu et la solidité des institutions. Tel est le trait distinctif de la voie empruntée : une modernité qui avance non pas contre le droit, mais par le droit, et qui trouve dans l'État de droit, la proportionnalité et la responsabilité les moyens de garantir que le progrès technologique demeure, à tout moment, irréductiblement Humain.

(DE): Die jüngste Entwicklung des Regulierungsrahmens der Europäischen Union in den Bereichen Daten, künstliche Intelligenz, Cybersicherheit und digitale Identität offenbart einen Prozess tiefgreifender institutioneller Transformation, der weit über die technische Dimension von Innovation hinausgeht und die Züge einer echten verfassungsrechtlichen Neupositionierung der europäischen Rechtsordnung annimmt. Die Digitalisierung ist dabei alles andere als ein bloß instrumentelles Phänomen; sie wird vielmehr zum Boden, auf dem die Union ihre politische und rechtliche Identität bekräftigt, indem sie ein *Governance*-Modell definiert, das technologischen Fortschritt, die Integrität der Institutionen und den substantiellen Schutz der menschlichen Person miteinander zu verbinden vermag. Das Ergebnis ist ein Innovationsrecht, das nicht am Rand der grundlegenden Architektur der Rechtsordnung steht, sondern sich in ihrem Zentrum verortet, weil es auf der Qualität der Daten, der Transparenz der Algorithmen und der öffentlichen Verantwortlichkeit als unverzichtbarer Voraussetzung demokratischer Legitimität aufbaut.
Die Gesamtheit der jüngsten Maßnahmen der Europäischen Union bildet ein Projekt von großer Tragweite, in dem technologische Innovation, regulatorische Effizienz und der Schutz des Einzelnen nicht in ein dialektisches, geschweige denn konfliktives Verhältnis gesetzt, sondern in eine einheitliche Konzeption digitaler Entwicklung als eines intrinsisch wertegeleiteten Prozesses integriert werden. Europa verpflichtet sich damit, ein Modell zu gestalten, in dem Technologie nicht als Faktor der Einschränkung von Freiheiten wirkt, sondern als Instrument, durch das Würde, Gleichheit und demokratische Teilhabe umfassender verwirklicht werden können. In diesem Sinne erscheint die Digitalisierung als der Ort, an dem sich die Fähigkeit der europäischen Rechtsordnung misst, ihre Zukunft zu steuern, und dabei zugleich die Zentralität des Individuums und die Solidität der Institutionen zu bewahren. Dies ist das kennzeichnende Merkmal des eingeschlagenen Weges: eine Modernität, die nicht gegen das Recht, sondern durch das Recht voranschreitet und die im Rechtsstaat, in der Verhältnismäßigkeit und in der Verantwortlichkeit die Mittel findet, um sicherzustellen, dass technologischer Fortschritt jederzeit irreduzibel menschlich bleibt.

(PT): A recente evolução do quadro regulatório da União Europeia em matéria de dados, inteligência artificial, cibersegurança e identidade digital revela um processo de profunda transformação institucional, que vai muito além da dimensão técnica da inovação para assumir as características de um verdadeiro reposicionamento constitucional da ordem jurídica europeia. A digitalização, longe de ser um fenómeno meramente instrumental, torna-se o terreno sobre o qual a União reafirma a sua identidade política e jurídica, definindo um modelo de governação capaz de conciliar o progresso tecnológico, a integridade das instituições e a proteção substancial da pessoa humana. O resultado é um direito da inovação que não é periférico à arquitetura fundamental da ordem jurídica, mas se situa no seu próprio centro, pois assenta na qualidade dos dados, na transparência dos algoritmos e na responsabilidade pública como condição indispensável da legitimidade democrática.
O conjunto das recentes intervenções da União Europeia constitui um projeto de amplo alcance, no qual a inovação tecnológica, a eficiência regulatória e a proteção do indivíduo não são colocadas numa relação dialética nem, muito menos, conflituosa, mas antes integradas numa conceção unificada do desenvolvimento digital como um processo intrinsecamente orientado por valores. A Europa compromete-se, assim, a construir um modelo em que a tecnologia não opere como um fator de restrição das liberdades, mas como um instrumento por meio do qual se realizem mais plenamente a dignidade, a igualdade e a participação democrática. Nesse sentido, a digitalização configura-se como o espaço em que se mede a capacidade da ordem jurídica

europeia para governar o seu futuro, preservando simultaneamente a centralidade do indivíduo e a solidez das instituições. Esta é a característica distintiva do caminho adotado: uma modernidade que avança não contra o direito, mas através do direito, e que encontra no Estado de direito, na proporcionalidade e na responsabilidade os meios para garantir que o progresso tecnológico permaneça, em todos os momentos, irreductivelmente Humano.

(IT): L'evoluzione recente del quadro regolatorio dell'Unione europea in materia di dati, intelligenza artificiale, sicurezza informatica e identità digitale rivela un processo di profonda trasformazione istituzionale che va ben oltre la dimensione tecnica dell'innovazione per assumere i tratti di un vero e proprio consolidamento giuridico attraverso un modello di governance capace di coniugare progresso tecnologico, assetto istituzionale e tutela sostanziale della persona. Ne emerge un diritto dell'innovazione che non è periferico rispetto all'architettura fondamentale dell'ordinamento, ma si colloca al suo stesso centro, poiché costruito sulla qualità dei dati, sulla trasparenza degli algoritmi e sulla responsabilità pubblica quale condizione imprescindibile di legittimazione democratica.
L'insieme dei recenti interventi dell'Unione europea compone un progetto di ampio respiro, nel quale l'innovazione tecnologica, l'efficienza regolatoria e la tutela della persona non sono collocate in rapporto dialettico e meno che mai conflittuale, ma integrate in una concezione unitaria dello sviluppo digitale come processo intrinsecamente valoriale. L'Europa si impegna così a costruire un modello nel quale la tecnologia non operi come fattore di compressione delle libertà, bensì come strumento attraverso il quale realizzare in modo più pieno la dignità, l'eguaglianza e la partecipazione democratica. In questo senso, la rete si configura come il luogo in cui si misura la capacità dell'ordinamento europeo di governare il proprio futuro, preservando al contempo la centralità della persona e la solidità delle istituzioni. È questa la cifra distintiva del percorso intrapreso, una modernità che avanza non contro il diritto, ma mediante il diritto, e che trova nella regola, nella proporzionalità e nella responsabilità gli strumenti attraverso cui assicurare che il progresso tecnologico rimanga sempre essenzialmente e irriducibilmente Umano.

Sommario: 1. La complessa costruzione del quadro regolatorio europeo e interno. – 2. Verso una *governance* basata sulla qualità delle informazioni, sull'interoperabilità delle infrastrutture e sulla certezza del diritto. – 3. L'equilibrio necessario tra innovazione e protezione dei diritti. – 4. La strategia regolatoria europea come strumento necessario per assicurare l'identità digitale e contrastare l'opacità algoritmica. – 5. Conclusioni.

1. La complessa costruzione del quadro regolatorio europeo e interno

L'ordinamento dell'Unione europea ha intrapreso un percorso di progressiva costruzione di un'architettura normativa in grado di orientare la trasformazione

digitale verso un equilibrio stabile tra innovazione tecnologica e tutela della persona umana. L'evoluzione dell'ecosistema digitale europeo si caratterizza così per una crescente consapevolezza circa la necessità di un assetto regolatorio capace di integrare in modo organico dati, intelligenza artificiale, cybersicurezza e identità digitale, collocando al centro un modello di governance fondato sulla responsabilità, sulla qualità delle informazioni e sulla garanzia della dignità umana. Tale approccio emerge con chiarezza nella Comunicazione della Commissione europea *Data Union* (*Strategy: Unlocking Data for Artificial Intelligence*, COM2025 835 final, 19 novembre 2025), secondo cui l'accessibilità e interoperabilità dei dati costituisce prerequisito essenziale per uno sviluppo sicuro e trasparente dei sistemi di intelligenza artificiale e, più in generale, condizione strutturale per generare valore economico e sociale nel rispetto dei diritti fondamentali.

La Commissione europea sottolinea inoltre l'esigenza di superare la frammentazione dell'*acquis* digitale, riducendo gli oneri amministrativi e accrescendo la prevedibilità normativa per cittadini e imprese. Questo orientamento è alla base della proposta di regolamento denominata *Digital Omnibus (European Commission, Proposal for a Regulation (Digital Omnibus), COM(2025) 837 final, 19 novembre 2025)*, che introduce misure di semplificazione riguardanti l'intelligenza artificiale, la cybersicurezza e la governance dei dati, eliminando sovrapposizioni normative e coordinando le fonti in un quadro più lineare e coerente. L'intervento si inserisce in un più ampio programma di "stress test" del *rulebook* digitale, volto a verificare la coerenza, l'efficacia e l'impatto cumulativo delle norme europee sul funzionamento del mercato unico [1]. La sem-

[1] G. De Minico, *Le fonti del diritto: un argine all'intelligenza artificiale?*, in *Riv. AIC*, 3/2025, p. 76 ss., ove viene evidenziato come «le fonti del diritto siano cambiate via via che le dinamiche di Internet e di intelligenza artificiale si evolvevano, presentando tre forme distinte che riflettono contesti storici, contemporanei e futuri. Inizialmente, l'analisi esplora la tradizionale autoregolazione, ricostruendone il percorso da semplice pratica autoregolativa a fonte del diritto, a condizione di rispettare i requisiti di legittimità posti dall'ordinamento europeo. Nel contesto contemporaneo, il lavoro approfondisce il ruolo dell'intelligenza artificiale, concentrandosi sui seguenti strumenti normativi individuati nell'A.I. Act: norme armonizzate e co-regolazione tra autorità pubbliche e privati. Queste fonti esemplificano una complessa interazione tra decisioni politiche e competenza tecnica, sollevando importanti questioni riguardo alla legittimità alla rappresentanza democratica e alla tutela dei diritti fondamentali. Guardando al futuro, l'Autrice discute il ruolo inedito della «tecnica come fonte del diritto», un modello che diverge radicalmente dai paradigmi tradizionali e autoregolati. Si vedano, inoltre, M. Tuozzo, *Note intorno al (futuro) Artificial Intelligence Act. Prospettive costituzionali e sfide dell'immigrazione nell'ecosistema digitale*, in *Dir. pubbl. eur. Rass. online*, 2024, n. spec. 1, p. 213 ss.; J. Black, *Rules and Regulations*, Oxford, 1997, pp. 71-80; A. Ogus, *Rethinking Self-Regulation*, in *Oxford JLS*, 15, 1995, pp. 97-108; O. Pollicino, *Regolazione e innovazione tecnologica nell'or-*

plificazione è considerata non come riduzione delle tutele, ma come rafforzamento della loro effettività: e ciò in quanto un *corpus* normativo più organico e leggibile aumenta la certezza del diritto e consente un'applicazione più uniforme delle garanzie.

È noto che l'evoluzione delle tecnologie emergenti tende a dilatare l'area delle interferenze tra diritto, tecnica e organizzazione pubblica, sollecitando un ripensamento del ruolo delle fonti e delle garanzie normative volte ad assicurare la tutela dei diritti fondamentali in un contesto dominato da processi decisionali automatizzati. In questa prospettiva, l'intervento regolatorio non può limitarsi a governare la tecnica, impedendo che diventi modello ordinante, ma soprattutto deve preservare la sovranità del diritto e con essa del principio democratico che trova il suo fondamento nel modello di bilanciamento dei diversi poteri dello Stato [2].

Ciò premesso, occorre ricordare che l'ordinamento digitale dell'Unione si articola attorno a tre principali regolamenti: il *Data Governance Act* (*Regulation EU 2022/868*) che istituisce meccanismi fiduciari di intermediazione e promuove l'altruismo dei dati; il *Data Act*, (*Regulation EU 2023/2854*), che disciplina i diritti di accesso e utilizzo dei dati generati da dispositivi connessi e rafforza l'interoperabilità dei servizi cloud; e l'*AI Act (Regulation EU 2024/1689*), che introduce un sistema multilivello basato sulla classificazione del rischio, sulla trasparenza algoritmica e sulla responsabilità degli operatori economici [3]. Si tratta di un mosaico regolatorio coerente con la visione europea di un ambiente digitale sicuro, interoperabile e *human-centric*, nel quale la tecnologia

dinamento della rete, in *Riv. AIC*, 2/2025, p. 152 ss.; G. DE GREGORIO, *The Normative Power of Artificial Intelligence*, in *Indiana JGLS*, 30, 2023, p. 73 ss.; E. CELESTE, *Digital Constitutionalism: A New Systematic Theorisation*, in *Int. RLCT*, 33, 2019, p. 76 ss.

[2] Osserva U. RUFFOLO, *L'*AI Act *e l'evoluzione dell'AI*, in U. RUFFOLO (a cura di), *AI Act. La regolamentazione europea dell'intelligenza artificiale*, Roma, 2025, p. 9, che l'*AI Act*, «considerato la prima e antesignana misura di sistematica regolazione dell'AI negli ordinamenti occidentali, spazio, tratta materia nella quale la velocità di evoluzione del fenomeno regolato può rischiare di rendere molti approcci regolatori strutturalmente inadeguati a segnare il passo. È il dilemma della dialettica regolante-regolato in un settore così "mobile" la cui natura e ritmo di crescita esigerebbero dalla *legal machinery*, secondo i massimi esponenti delle altre scienze ("dure" e non: da Bostrom ad Acemoglu), una sapienza regolatoria proattiva e non semplicemente reattiva».

[3] Cfr. in materia, G. FINOCCHIARO, *La proposta di regolamento sull'intelligenza artificiale: il modello europeo basato sulla gestione del rischio*, in *Dir. inf.*, 2022, p. 303 ss.; L. FLORIDI, *The European Legislation on AI: A Brief Analysis of its Philosophical Approach*, in *PhT*, 34, 2021, p. 215 ss.; G. MAZZINI-S. SCALZO, *The Proposal for the Artificial Intelligence Act: Considerations around Some Key Concepts*, in C. CAMARDI (a cura di), *La via europea per l'Intelligenza Artificiale*, Milano, 2023, p. 21 ss.; L. GATT-M.T. LO CONTE-M.G. MAZZARELLA, *L'ambito soggettivo ed oggettivo di applicazione dell'AI Act*, in U. RUFFOLO (a cura di), *op. cit*., p. 49.

sia governata da criteri di proporzionalità, documentazione e supervisione. L'*AI Act*, in particolare, definisce una struttura di responsabilità che integra requisiti tecnici e garanzie giuridiche, prevedendo un controllo umano significativo e standard stringenti di qualità e affidabilità dei processi di acquisizione e gestione delle informazioni [4].

A tali strumenti si affianca la proposta di regolamento sui Portafogli Aziendali Europei (*European Commission, Proposal for a Regulation on European Business Wallets, COM2025 838 final*), concepita per creare un'infrastruttura unionale interoperabile di identificazione, attestazione e scambio documentale tra imprese. Un ruolo determinante è attribuito ai *Data Labs* (*Supporting Infrastructure for High-Quality Data Spaces, in COM202) 835 final*), immaginati per offrire servizi fiduciari di preparazione, pseudonimizzazione e condivisione dei dati in ambienti sicuri. Essi costituiscono spazi controllati in cui maturano i processi di miglioramento della qualità dei dataset e nei quali vengono ridotte le distorsioni informative che potrebbero compromettere l'attendibilità dei sistemi di IA.

La riflessione più recente ha messo in luce come la qualità semantica e strutturale dei dati influenzi quella delle decisioni algoritmiche. In questo senso, è stato osservato [5] quanto sia elevato l'impatto del linguaggio dell'intelligenza artificiale sulla tutela dei diritti fondamentali, ben potendo l'opacità algoritmica

[4] A. ALAIMO, *Il Regolamento sull'Intelligenza Artificiale: dalla proposta della Commissione al testo approvato dal Parlamento. Ha ancora senso il pensiero pessimistico?*, in *federalismi.it*, 25, 2023, p. 133 ss., secondo cui «la valutazione dell'*AI Act* non deve rimanere isolata dal generale contesto normativo europeo e nazionale in cui la proposta è concepita, ma debba piuttosto inquadrarsi – per usare un'immagine – con un grandangolo, sì da tenere adeguatamente conto delle molte discipline vigenti e dei numerosi progetti normativi in corso. Ciò consente di intercettare i fili di una trama normativa unitaria che, in un'ottica complessiva di governo dell'ecosistema digitale, non sembra che stia del tutto confinando ai margini le questioni sociali rispetto agli imperativi economici. Il quadro complessivo lascia meno spazio al pessimismo iniziale e suggerisce di non confinare la valutazione del progetto di AI Act nel libro nero delle iniziative unicamente finalizzate a regolare il mercato unico».

[5] A. SIMONCINI, *Il linguaggio dell'intelligenza artificiale e la tutela costituzionale dei diritti*, in *Riv. AIC*, 2/2023, p. 8, per il quale «un numero crescente di decisioni che interferiscono con la nostra sfera di libertà sono prese avvalendosi di sistemi di IA (che, inevitabilmente, tendono a sostituire e non solo a supportare la decisione). L'evoluzione di questo specifico settore tecnologico sta portando alla sostituzione di modelli di inferenza causale con modelli logico-statistici basati sulla correlazione. Questa evoluzione rende in alcuni casi estremamente difficile, se non del tutto impossibile, comprendere il linguaggio e, dunque, le ragioni di tali decisioni. Alcuni, seppur rari, casi giurisprudenziali emblematici confermano/evidenziano i punti precedenti»; ID., *L'uso delle tecnologie nella pandemia e le nuove diseguaglianze*, in A. PAJNO-L. VIOLANTE (a cura di), *Biopolitica, pandemia e democrazia*. *Rule of law nella società digitale*, II, *Etica, comunicazione e diritti*, Bologna, 2021.

incidere sulla comprensione delle decisioni pubbliche e sulla trasparenza dell'esercizio del potere. Il problema del linguaggio dell'IA si collega alla necessità di elaborare un diritto alla comprensibilità tecnologica, capace di garantire che l'automazione non comprometta il diritto di difesa e la stessa effettività del principio di legalità.

In altri ambiti (diritto amministrativo, diritto tributario) [6], l'integrazione dei dati e l'impiego di strumenti predittivi stanno ridefinendo la relazione tra amministrazioni pubbliche (e quindi quelle fiscali) e cittadino-contribuente, ponendo in una nuova luce il difficile equilibrio tra efficienza, legalità e capacità contributiva [7]. Da una prospettiva più ampia [8], l'utilizzo delle tecnologie digitali

[6] G. SALANITRO, *Una prima lettura sull'impatto della legge IA nel diritto tributario*, in *Riv. tel. dir. trib.*, 9 ottobre 2025; L. TOSI, *L'evoluzione digitale del sistema tributario*, in *Tax News*, 1/2025, p. 32 ss. Per approfondimenti, R. CORDEIRO GUERRA, *L'intelligenza artificiale nel prisma del diritto tributario*, in *Dir. prat. trib.*, 2020, I, p. 922 ss., secondo cui «quando si affaccia al mondo del diritto un nuovo fenomeno o una nuova forma, anche embrionale, di entità prospetticamente configurabile quale centro di imputazione di diritti ed obblighi, quello tributario è tra i primi o comunque più solerti settori del sapere giuridico a scandagliarne le potenzialità, ovviamente nella prospettiva di dar vita ad un nuovo soggetto passivo da sottoporre a tassazione o al quale addossare obblighi strumentali al prelievo. Tale dinamica si sta ripetendo anche rispetto al fenomeno dell'intelligenza artificiale, per un verso con fughe in avanti costruite su una futura possibilità di tassare macchine capaci di autodeterminarsi e dunque anche di produrre ricchezza loro imputabile; e per l'altro, più realisticamente, con riferimento al trattamento fiscale da riservare ad imprese o privati che utilizzano forme di intelligenza artificiale in grado di sostituire, in tutto o in parte, il lavoro umano. In una diversa e ulteriore prospettiva, altrettanto fervida di sviluppo, l'interesse verso l'intelligenza artificiale si è poi appuntato sul potenziale impatto che essa potrà avere nell'implementazione e applicazione delle disposizioni tributarie: uno scenario al tempo stesso fecondo e sconfinato, considerata la mole delle disposizioni, la loro crescente complessità, il numero di soggetti coinvolti nella loro quotidiana attuazione negli attuali sistemi di fiscalità di massa. In quest'ultima chiave, in particolare, spiccano le potenzialità, ma anche le problematiche, che l'uso dell'intelligenza artificiale potrebbe comportare ai fini dei controlli sui contribuenti, orientando e rendendo più neutrale la scelta dei soggetti da sottoporre a verifica». Si veda anche A. FIDELANGELI-F. GALLI, *The Use of AI Technologies in the Tax Law Domain: An Interdisciplinary Analysis*, in *Dir. prat. trib. int.*, 2022, pp. 125-126; S.A. PARENTE, *La soggettività tributaria delle macchine intelligenti e i nuovi modelli di tassazione dei processi produttivi automatizzati e interconnessi*, in *Ianus-Diritto e Finanza*, 23, 2021, pp. 111-112.

[7] Cfr. S. DORIGO, *The 'Algorithmic Revolution': Fair Taxation, Social Pact and Global Governance*, in M. BELOV (a cura di), *The IT Revolution and its Impact on State, Constitutionalism and Public Law*, Oxford, 2021, p. 161 ss.; ID., *Intelligenza artificiale e norme antiabuso: il ruolo dei sistemi 'intelligenti' tra funzione amministrativa e attività giurisdizionale*, in S. DORIGO (a cura di), *Il ragionamento giuridico nell'era dell'intelligenza artificiale*, Pisa, 2020, p. 123.

[8] Per ulteriori approfondimenti sui profili tributari dell'economia digitale, si consenta il rinvio al volume F. GALLO-A.F. URICCHIO, *La tassazione dell'economia digitale*, Bari, 2022; A.F. URICCHIO, *Evoluzione tecnologica e imposizione: la cosiddetta Bit Tax. Prospettive di riforma*

nei processi produttivi come nella vita quotidiana dà luogo ad una profonda trasformazione degli stili di vita e quindi del modo di lavorare, consumare e generare ricchezza. La transizione digitale necessita quindi di un forte sforzo tecnologico, infrastrutturale, amministrativo e persino regolatorio [9]. Dopo essere passati da semplici *consumer* dei servizi digitali a *prosumer*, in grado contestualmente di essere fruitori e produttori dei contenuti generati, siamo di fronte a un'ulteriore evoluzione in cui ciascuno di noi ha preso in carico una parte significativa della produzione dai dati nella società neo-industriale, acquisendo una nuova consapevolezza del loro valore. La *data economy* pone poi un ulteriore tema, ad oggi, poco esplorato: la generazione del dato, spesso gratuita e spontanea, diviene in capo all'utilizzatore professionale valore che però resta in capo allo stesso senza alcuna restituzione all'utente né all'intera collettività attraverso gli strumenti della fiscalità generale.

Evidente è che la costruzione dell'ecosistema digitale europeo necessita di un quadro istituzionale che assuma l'innovazione non come accelerazione incontrollata ma come processo ordinato [10] e orientato al benessere collettivo, alla

della fiscalità di internet, in *Dir. inf.*, 2005, p. 753; F. GALLO, *Fisco ed economia digitale*, in *Dir. prat. trib.*, 2015, I, p. 601; ID., *Il futuro non è un vicolo cieco. Lo Stato tra globalizzazione, decentramento ed economia digitale*, Palermo, 2019; L. DEL FEDERICO-C. RICCI (a cura di), *La Digital Economy nel sistema tributario italiano ed europeo*, Padova, 2015; L. DEL FEDERICO (a cura di), *Le nuove forme di tassazione della Digital Economy*, Roma, 2017; ID., *La tassazione nell'era digitale. Genesi, diffusione ed evoluzione dell'equalisation levy*, in *Dir. prat. trib. int.*, 2020, p. 1431; L. CARPENTIERI (a cura di), *Profili fiscali dell'economia digitale*, Torino, 2020; R. CORDEIRO GUERRA-S. DORIGO (a cura di), *Fiscalità dell'economia digitale*, Pisa, 2022; A. DI PIETRO-P. SANTIN (a cura di), *La fiscalità dell'economia digitale tra Italia e Spagna*, Padova, 2021; F. FARRI, *Tax Sovereignty and the Law in the Digital and Global Economy*, Torino, 2021; S. DORIGO (a cura di), *Il ragionamento giuridico nell'era dell'intelligenza artificiale*, Pisa, 2020; A. PERSIANI (a cura di), *La tassazione dell'economia digitale tra sviluppi recenti e prospettive future*, Atti del convegno del 29 novembre 2018 (numero speciale di Diritto, mercato, tecnologia), *NEU,* 2019; G. CORASANITI, *La tassazione della "digital economy": evoluzione del dibattito internazionale e prospettive nazionali*, in *Dir. prat. trib. int.*, 2020, I, p. 1397; M.C. FREGNI, *Mercato digitale e tassazione: misure attuali e progetti di riforma*, in *Riv. dir. fin.*, 2017, I, p. 51; C. BUCCICO, *Problematiche e prospettive della tassazione dell'economia digitale*, in *Dir. proc. trib.*, 2019, p. 255; FONDAZIONE BRUNO VISENTINI, *La tassazione dell'economia digitale*, 4 aprile 2019, 476; E. DELLA VALLE-G. FRANSONI (a cura di), *L'imposta sui servizi digitali*, Padova, 2021.

[9] B.-J. KOOPS, *Ten Dimensions of Technology Regulation*, in M.E.A. GOODWIN-B.-J. KOOPS-R.E. LEENES (a cura di), *Dimensions of Technology Regulation*, Nijmegen, 2010, p. 309 ss.

[10] Cfr. F. GALLO, *Cyberspazio, diritti e tutele*, in A.F. URICCHIO-C. CALDAROLA (a cura di), *L'intelligenza artificiale tra regolazione e esperienze applicative. AI ACT (regolamento UE) e disegno di legge governativo*, Bari, 2025, p. 16, il quale sottolinea la necessità, per le istituzioni europee, di definire un quadro regolatorio chiaro in materia di intelligenza artificiale, osservando come, se da un lato viviamo ormai in una sorta di metaverso popolato da oggetti "intelligenti" che ci parlano, ci ascoltano e ci assistono nelle più varie attività quotidiane, dall'altro si facciano

sostenibilità sociale e al pieno rispetto della dignità della persona umana. In questo contesto, l'Italia si è distinta per la celerità con cui ha recepito l'impianto regolatorio europeo in materia di intelligenza artificiale; di recente è stata approvata definitivamente la legge 23 settembre 2025, n. 132, disposizioni e deleghe al Governo in materia di intelligenza artificiale [11] che, prima in Europa, affianca al regolamento una compiuta disciplina nazionale, introducendo elementi di specificità, tra cui il riferimento espresso ad un modello antropocentrico ed eticamente orientato, la promozione della ricerca attraverso la cooperazione tra università, enti pubblici e imprese, l'estensione della disciplina a settori come la pubblica amministrazione e la disabilità. La nuova disciplina non si limita ad accogliere i principi cardine enunciati dall'Unione ma li rafforza, intervenendo in modo puntuale e vincolante nelle parti in cui l'ordinamento eurounionale lasciava margini di discrezionalità. Valori come il rispetto della persona e della sua dignità, delle libertà personali e dei diritti fondamentali sono pienamente affermati. La legge nazionale dedica, altresì, un'attenzione specifica a settori "sensibili", particolarmente importanti e delicati, come la pubblica amministrazione, la salute e la disabilità – quest'ultima non menzionata nel Regolamento Europeo. Si tratta di novità meritorie e di grande rilievo, destinate non solo a incidere sulla regolazione giuridica, ma anche a orientare lo sviluppo della ricerca scientifica e dell'evoluzione tecnologica. Occorre precisare che né il Regolamento europeo né la legge italiana adottano un approccio repressivo o soffocante nei confronti dell'innovazione tecnologica. Al contrario, la consapevolezza sottesa è che la normativa possa costituire un efficace strumento di promozione di un utilizzo consapevole e responsabile dell'intelligenza artificiale. Come ricorda un antico proverbio cinese, «non si può arrestare un fiume in piena con la mano, ma si può costruire un argine». Il compito del legislatore è stato dunque proprio quello di delineare tali argini, affinché l'intelligenza artificiale – vero e proprio fiume impetuoso – non esondi, ma possa essere indirizzata a generare crescita, sviluppo ed equità per il nostro Paese.

Nella legge italiana, significativo è l'art. 3, rubricato «Principi generali», secondo cui lo sviluppo, l'adozione e l'utilizzo di sistemi e modelli di intelligenza artificiale a finalità generali debbono avvenire nel rispetto dei diritti fondamentali, delle libertà costituzionali e del diritto dell'Unione europea, nonché secondo i principi di trasparenza, proporzionalità, sicurezza, protezione dei dati personali,

sempre più pressanti i dubbi circa le possibili ombre della digitalizzazione, la natura "artificiale" dei suoi frutti, i rischi di manipolazione, l'assenza di adeguati controlli e il potenziale decremento occupazionale che essa potrebbe produrre. Da qui la necessità di una normativa europea idonea a prevenire, nel tempo, tali effetti.

[11] In GU Serie Generale n. 223 del 25 settembre 2025, n. 25G00143.

riservatezza, accuratezza, non discriminazione, parità di genere e sostenibilità (primo comma). Viene, in tal modo, delineata una declinazione dell'intelligenza artificiale orientata alle finalità e agli interessi generali, secondo un approccio improntato al principio di proporzionalità e fondato su una rigorosa valutazione del rischio. In tale prospettiva, la sicurezza digitale assume il rango di requisito inderogabile, sorretto da controlli mirati e stringenti, destinati ad assicurare la necessaria resilienza del sistema rispetto a possibili alterazioni, manipolazioni o usi distorti delle tecnologie. Particolarmente significativa è la previsione del comma 7 dello stesso articolo che consacra il diritto delle persone con disabilità ad un accesso pieno ed effettivamente non discriminatorio ai sistemi di IA e alle relative funzionalità, in conformità alla Convenzione ONU sui diritti delle persone con disabilità del 2006, ratificata in Italia con legge n. 18/2009; si tratta di una formula che inserisce la tecnologia entro l'orizzonte della dignità e dell'uguaglianza sostanziale. Nell'articolato normativo, il Capo II si presenta di cruciale importanza, enunciando, in maniera puntuale e sistematica, principi generali e regole specifiche per l'impiego dell'intelligenza artificiale all'interno di settori particolarmente sensibili, ove l'incidenza delle decisioni algoritmiche risulta concretamente idonea a produrre conseguenze rilevanti e, talvolta, irreversibili sui diritti fondamentali della persona. In tale quadro, il legislatore, con l'art. 7, ha espressamente e rigorosamente vietato l'utilizzo di modelli algoritmici fondati su criteri manifestamente discriminatori, i quali potrebbero inopinatamente condizionare o ingiustamente limitare l'accesso alle prestazioni sanitarie, configurandosi così quale presidio essenziale e inderogabile a tutela dell'universalità, concreta ed effettiva, delle cure. In modo altrettanto significativo, la modifica apportata all'art. 9 c.p.c. si inserisce coerentemente in una prospettiva di certezza e di garanzia, attribuendo con chiarezza e senza ambiguità al tribunale la competenza a conoscere delle controversie relative al funzionamento dei sistemi di intelligenza artificiale, con l'effetto di dissipare definitivamente ogni incertezza interpretativa circa l'individuazione del giudice competente.

Accanto agli aspetti strettamente regolatori, la legge introduce altresì precisi e significativi indirizzi di politica attiva. In particolare, l'art. 5 affida espressamente allo Stato il compito di promuovere, in maniera sistematica e coerente, lo sviluppo e l'impiego dell'intelligenza artificiale quale strumento privilegiato volto a migliorare l'interazione uomo-macchina, anche attraverso l'impiego avanzato della robotica, favorendo contestualmente la ricerca scientifica e incoraggiando la cooperazione virtuosa tra università e imprese. Tali coordinate si configurano come snodi essenziali del rapporto tra sapere accademico e tessuto produttivo, destinati non soltanto a rafforzare stabilmente la competitività del Paese, ma anche a catalizzare investimenti strategici nei settori nevralgici dell'economia nazionale, entro un quadro organico e dinamico

che intreccia armoniosamente innovazione, crescita sostenibile e superiore interesse pubblico.

2. Verso una *governance* basata sulla qualità delle informazioni, sull'interoperabilità delle infrastrutture e sulla certezza del diritto

L'evoluzione del quadro regolatorio dell'Unione europea in materia di dati, intelligenza artificiale, *cybersicurezza* e identità digitale rivela un processo di consolidamento progressivo volto a realizzare un sistema disciplinare unitario, leggibile e capace di sostenere lo sviluppo tecnologico nel rispetto dei valori fondativi dell'ordinamento. L'architettura digitale dell'Unione, pur articolata su fonti settoriali eterogenee, tende progressivamente ad assumere la forma di un vero e proprio "diritto dell'integrazione tecnologica", nel quale ciascun regolamento non opera in funzione isolata ma contribuisce alla costruzione di una cornice unitaria. Tale evoluzione rivela una trasformazione del ruolo del legislatore europeo, chiamato non solo a disciplinare nuovi mercati, ma anche a presidiare la qualità sistemica delle interazioni digitali tra istituzioni, cittadini e operatori economici. In questa prospettiva, la *governance* dei dati e dell'intelligenza artificiale non si configura come un insieme di atti tecnici, bensì come un processo di ingegneria istituzionale destinato a incidere sulla struttura stessa del mercato unico e sulla configurazione dei poteri pubblici nell'era algoritmica. Tale processo si inserisce in una strategia istituzionale ampia, nella quale la Commissione europea ha progressivamente delineato un modello di governance basato sulla qualità delle informazioni, sull'interoperabilità delle infrastrutture e sulla certezza del diritto, riconoscendo che l'affidabilità dei sistemi digitali costituisce presupposto indispensabile per il buon funzionamento del mercato unico. La *Data Union Strategy* sottolinea che la frammentazione normativa rappresenta uno dei principali ostacoli alla maturazione degli ecosistemi digitali e che, pertanto, la costruzione di un'unica architettura deve procedere attraverso un'armonizzazione delle discipline, un coordinamento delle fonti e una semplificazione delle procedure amministrative [12]. La strategia attribuisce alla qualità dei dati un valore sistemico, evidenziando come la loro disponibilità in forma interoperabile costituisca condizione essenziale per l'adozione sicura e responsabile dell'intelligenza artificiale e per la diffusione di modelli di innovazione inclusivi e sostenibili.

[12] European Commission, *Data Union Strategy: Unlocking Data for Artificial Intelligence*, cit.; sul valore sistemico della qualità dell'informazione e della governance dei dati nel quadro europeo v. F. GALLO, *Cyberspazio, diritti e tutele*, cit., p. 9, ove la qualità dei dati è considerata presupposto della tutela dei diritti e della legittimazione democratica dei processi digitali.

Il primo pilastro di tale architettura è rappresentato dal *Data Governance Act* [13], che introduce un modello istituzionale fondato sulla fiducia nella condivisione dei dati. Il regolamento istituisce i servizi di intermediazione, concepiti come soggetti neutrali deputati a facilitare la cessione volontaria e controllata dei dati tra detentori e utilizzatori e disciplina l'altruismo dei dati, configurato come messa a disposizione volontaria e non remunerata di informazioni per finalità di interesse generale. L'atto definisce, inoltre, procedure uniformi per il riutilizzo delle informazioni detenute dal settore pubblico, rafforzando la trasparenza amministrativa e ponendo le basi per la creazione di spazi europei dei dati affidabili e interoperabili. La dimensione fiduciaria delle informazioni, più volte richiamata dalla dottrina, si collega ai profili relativi alla responsabilità pubblica nella gestione delle infrastrutture digitali e al ruolo delle autorità nazionali nel garantire che la circolazione dei dati avvenga in condizioni di sicurezza, trasparenza e certezza giuridica [14].

Il secondo pilastro è costituito dal *Data Act* [15], che definisce diritti e obblighi relativi all'accesso ai dati generati da dispositivi connessi e servizi digitali. Il regolamento riconosce all'utente la facoltà di accedere ai dati prodotti dall'oggetto e impone obblighi di trasparenza, prevedibilità contrattuale e interoperabilità. La disciplina si estende ai servizi *cloud* ed *edge*, imponendo standard tecnici finalizzati a garantire la portabilità tra fornitori, ridurre i rischi di *lock-in* tecnologico e favorire la mobilità dei dati. Interoperabilità, qualità dei dati e certezza delle condizioni contrattuali costituiscono condizioni indispensabili per un mercato equo e competitivo e, in particolare, per un rapporto equilibrato tra operatori economici, cittadini e amministrazioni pubbliche [16].

---

[13] Reg. (UE) 2022/868 (*Data Governance Act*); sulla dimensione istituzionale della condivisione dei dati e sul ruolo della *governance* pubblica v. A.F. URICCHIO, *Il fisco dinanzi all'intelligenza artificiale e alla digital transformation*, in A.F. URICCHIO-C. CALDAROLA (a cura di), *op. cit*., p. 197 ss., che qualifica la *governance* dei dati come infrastruttura ordinante dell'ordinamento digitale.

[14] Sul nesso tra fiducia, qualità del dato e responsabilità pubblica D.U. GALLETTA-B. CAPPELLO, *Il Regolamento UE sull'intelligenza artificiale e le amministrazioni pubbliche*, in A.F. URICCHIO-C. CALDAROLA (a cura di), *op. cit*., p. 59 ss., in relazione ai rischi di opacità e alla necessità di trasparenza algoritmica.

[15] Reg. (UE) 2023/2854 (*Data Act*); per una lettura civilistica degli effetti dell'accesso ai dati e dell'interoperabilità sulle categorie tradizionali cfr. U. RUFFOLO, *Intelligenza artificiale e responsabilità: servono (sempre) regole nuove per il nuovo fenomeno?*, in A.F. URICCHIO-C. CALDAROLA (a cura di), *op. cit*., p. 93 ss.

[16] Sul rapporto tra interoperabilità, certezza giuridica e funzionamento equo del mercato, con particolare riferimento all'ambito tributario e amministrativo, si rinvia a A.F. URICCHIO, *Il fisco dinanzi all'intelligenza artificiale e alla digital transformation*, cit., p. 201 ss., ove la qualità dei dati è indicata quale condizione della sostenibilità dei sistemi pubblici digitalizzati.

Il terzo pilastro è rappresentato dall'*AI Act* [17], primo regolamento organico dell'Unione europea dedicato all'intelligenza artificiale. Il testo introduce un modello di rischio graduato e differenziato in sistemi vietati, sistemi ad alto rischio e sistemi soggetti a obblighi di trasparenza, delineando requisiti tecnici, obblighi documentali e procedure di conformità differenziate. Il regolamento attribuisce alla persona umana il ruolo di supervisore attivo dei sistemi ad alto rischio, riconoscendo che, anche in presenza di processi automatizzati complessi, l'intervento umano deve rimanere il perno del controllo democratico. E ciò in quanto la sostenibilità dell'intelligenza artificiale dipende dalla capacità dell'ordinamento di assicurare un equilibrio armonico tra autonomia tecnica, responsabilità degli operatori e garanzie sostanziali e procedurali.

Un ruolo significativo nella costruzione dell'ecosistema digitale è svolto dalle iniziative di semplificazione normativa adottate dalla Commissione europea. Il *Digital Omnibus* [18] propone un'integrazione sistemica degli atti settoriali in materia di dati, intelligenza artificiale e cybersicurezza, ricondotti in un *corpus* unificato. La proposta introduce adattamenti tecnici destinati a garantire coordinamento, chiarezza e applicazione uniforme, oltre a prevedere l'istituzione di un punto unico europeo di notifica degli incidenti di cybersicurezza, finalizzato a ridurre la duplicazione degli obblighi informativi e a migliorare il coordinamento tra autorità nazionali ed enti europei [19]. L'intervento è parte di una più ampia azione di verifica della coerenza del quadro digitale unionale, con l'obiettivo di ridurre la complessità regolatoria e rafforzare la prevedibilità del diritto.

La disciplina dei *Portafogli Aziendali Europei* [20] completa il quadro, istituendo un'infrastruttura unitaria per la gestione dell'identità digitale delle imprese e per la condivisione transfrontaliera delle attestazioni aventi valore giuridico. Tale strumento risponde all'esigenza, rilevata dalla dottrina economica

[17] Reg. (UE) 2024/1689 (*AI Act*); per una ricostruzione sistematica dell'architettura del rischio e degli effetti sull'azione amministrativa si vedano C. CALDAROLA, *L'AI Act: il nuovo regolamento europeo sull'intelligenza artificiale*, in A.F. URICCHIO-C. CALDAROLA (a cura di), *op. cit*., p. 87 ss.; M. CORRADINO, *Potere pubblico e intelligenza artificiale*, *ivi*, p. 39 ss.

[18] European Commission, *Proposal for a Regulation on Artificial Intelligence, Cybersecurity and Data Governance (Digital Omnibus)*, COM (2025) 837 final; sulla semplificazione normativa e sull'esigenza di coerenza sistemica del diritto digitale europeo v. A.F. URICCHIO, *Postfazione*, in A.F. URICCHIO-C. CALDAROLA (a cura di), *op. cit*., p. 467 ss.

[19] European Commission, *SWD (2025) 836 final – Digital Simplification Agenda*; Reg. (UE) 2024/1183 sul quadro europeo di identità digitale; sul ruolo delle infrastrutture fiduciarie e della verificabilità delle informazioni nelle relazioni economiche digitali. Cfr. J.D. LOHR-W.J. MAXWELL-P. WATTS, *Legal Practitioners' Approach to Regulating AI Risks*, in K. YEUNG-M. LODGE (a cura di), *Algorithmic Regulation*, Oxford, 2019, p. 242 ss.

[20] Reg. (UE) 2024/1183.

e amministrativa, di predisporre un ambiente affidabile per le interazioni digitali, riducendo oneri burocratici, rafforzando la verificabilità delle informazioni e sostenendo l'integrazione del mercato interno [21].

L'evoluzione dell'*acquis* digitale non può essere interpretata come una mera ricomposizione tecnica delle fonti, poiché essa è strettamente connessa alla dimensione assiologica del diritto europeo, esprimendo il rapporto tra regolazione, tecnica e valori costituzionali nel quale il diritto non subisce la tecnica, ma la governa, indirizzando l'innovazione verso la tutela della persona e la salvaguardia dei principi democratici [22]. Alla luce di tali considerazioni, la costruzione dell'ecosistema digitale europeo costituisce una componente essenziale del processo di integrazione geopolitica [23], poiché la trasformazione tecnologica modifica non soltanto i rapporti economici e le categorie giuridiche fondamentali ma anche i rapporti tra Stati. Il diritto europeo, attraverso il coordinamento tra dati, identità digitale e intelligenza artificiale, tende, quindi, progressivamente a definire una nuova architettura istituzionale orientata alla trasparenza, all'affidabilità e alla sostenibilità in grado di incidere anche su scenari geopolitici internazionali.

3. L'equilibrio necessario tra innovazione e protezione dei diritti

Come evidenziato, la politica digitale dell'Unione europea si orienta verso un modello nel quale lo sviluppo tecnologico e la protezione dei diritti fonda-

[21] Cfr. V.S. POLI, *Il rafforzamento della sovranità tecnologica europea e il problema delle basi giuridiche*, in *Quaderni AISDUE*, 5/2021; V.M. INGLESE, *Il Regolamento sull'intelligenza artificiale come atto per il completamento e il buon funzionamento del mercato interno?* in *Quaderni AISDUE*, 2/2024; A. ADINOLFI, *L'intelligenza artificiale tra rischi di violazione dei diritti fondamentali e sostegno alla loro promozione: considerazioni sulla (difficile) costruzione di un quadro normativo dell'Unione*, in A. PAJNO-F. DONATI-A. PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, Bologna, 2022, p. 133 ss.; G. MOBILIO, *L'intelligenza artificiale e i rischi di una 'disruption' della regolamentazione giuridica*, in *Bio LJ*, 2020, p. 401 ss.

[22] Cfr. L. MEZZETTI-S. ROMANO, *Intelligenza artificiale, Costituzione, libertà*, in U. RUFFOLO (a cura di), *AI Act. La regolamentazione europea dell'intelligenza artificiale*, cit., p. 35, secondo cui «la dimensione dinamica che sospinge l'innovazione tecnologica si infrange sullo schermo statico della tutela dei principi e dei valori fondamentali che dovrebbero riverberarsi in tecniche di normazione che tengano conto dei principi-guida per un design etico della rivoluzione digitale».

[23] Cfr. F. GIANCOTTI, *Strategia e strategie IA: manca quella militare*, in A.F. URICCHIO-C. CALDAROLA (a cura di), *op. cit*., p. 355 ss., sul ruolo dell'intelligenza artificiale nella ridefinizione degli equilibri geopolitici e della sovranità digitale, e anche M. CALIGIURI, *L'intelligenza artificiale tra regolamentazione europea e strategie nazionali*, *ivi*, p. 475 ss.

mentali non si contrappongono, ma si integrano, poiché l'equità dei processi decisionali, la qualità dei dati e l'affidabilità dei sistemi costituiscono condizioni necessarie per un'innovazione realmente orientata al progresso collettivo [24]. La *Data Union Strategy* intende assicurare la costruzione di un ecosistema digitale sostenibile rispettoso della dignità della persona e coerente con i principi della Carta dei diritti fondamentali dell'Unione [25]. In tale prospettiva, la Commissione individua nella "governance responsabile" un elemento cardine per prevenire derive tecnologiche che potrebbero compromettere la fiducia degli utenti e la legittimazione democratica dei processi digitali.

L'Unione europea delinea un modello regolatorio nel quale la tutela dei diritti fondamentali non si colloca in una fase meramente successiva ai processi tecnologici, assumendo, invece, una funzione pienamente propulsiva, orientando la progettazione stessa degli ecosistemi digitali. Il principio di dignità, la garanzia contro ogni forma di discriminazione, il diritto a una partecipazione informata, e sotto certi aspetti persino il diritto alla verità [26], non operano, pertanto, come limiti esterni o freni all'innovazione e costituiscono parametri sostanziali destinati a guidare l'intero ciclo di vita delle tecnologie, dall'ideazione alla messa in esercizio [27]. Ne deriva una trasformazione significativa del rap-

[24] Sul tema, si veda J.R. REIDENBERG, *Lex Informatica: The Formulation of Information Policy Rules through Technology*, in *Texas LR*, 76, 1997, p. 555, il quale evidenzia come «le capacità tecnologiche e le scelte di progettazione dei sistemi impongano regole ai partecipanti». Tale approccio è stato successivamente ripreso e sistematizzato da L. LESSIG, *Code Is Law*, in *Harvard M*, 1° gennaio 2000, secondo cui il codice informatico si configura quale forma di "legge", in quanto idoneo a incidere sulle libertà individuali e a conformare l'esercizio dei diritti nello spazio digitale.

[25] European Commission, *Data Union Strategy: Unlocking Data for Artificial Intelligence*, COM(2025) 835 final, 19 novembre 2025; sul rapporto tra dignità della persona, qualità dei dati e legittimazione democratica dei processi digitali.

[26] Cfr. C.M. REALE-M. TOMMASI, *Libertà di espressione, nuovi media e intelligenza artificiale: la ricerca di un nuovo equilibrio nel nuovo ecosistema costituzionale*, in *DPCE online*, 1/2022, p. 326 ss. Gli Autori mettono in evidenza come l'innovazione tecnologica, e in particolare l'uso dell'IA nei processi comunicativi, imponga una rilettura dei tradizionali parametri costituzionali, richiedendo di contemperare l'espansione delle possibilità espressive con le nuove minacce derivanti da fenomeni quali la disinformazione algoritmica e l'opacità decisionale dei sistemi automatizzati.

[27] Sul carattere trasformativo della società per effetto della tecnologia si veda L. FLORIDI-F. CABITZA, *L'intelligenza artificiale: l'uso delle nuove macchine*, Milano, 2021, p. 9, evidenziano come «l'intelligenza artificiale è una delle forze più trasformative del nostro tempo ed è destinata a modificare profondamente il tessuto della società». Si veda anche A.F. URICCHIO-G. RICCIO-U. RUFFOLO (a cura di), *Intelligenza Artificiale tra etica e diritti. Prime riflessioni a seguito del libro bianco dell'Unione europea*, Bari, 2020, ove viene posto in risalto il bisogno di un approccio regolatorio ed etico in grado di accompagnare lo sviluppo dell'IA, scongiurando derive po-

porto tradizionale tra diritto e tecnica, poiché non è più il diritto a inseguire l'evoluzione tecnologica, ma è l'innovazione a essere chiamata a svilupparsi entro un quadro normativo definito e non negoziabile in quanto ispirato da valori condivisi e principi fondamentali propri della struttura dell'ordinamento europeo e di quelli costituzionali degli stati membri. Il principio di responsabilità tecnologica si traduce nell'esigenza di assicurare che l'automazione rimanga comprensibile, contestualizzata e verificabile, affinché la complessità algoritmica non generi spazi opachi tali da eludere il controllo umano o ridurre la capacità di verificare la coerenza delle decisioni con i principi del diritto.

La Commissione europea evidenzia che uno dei principali nodi interpretativi riguarda la disponibilità di dati idonei a supportare l'addestramento dei sistemi di intelligenza artificiale. La scarsità di *dataset* adeguati, accurati e rappresentativi costituisce un ostacolo significativo alla maturazione degli algoritmi e richiede infrastrutture fiduciarie in grado di garantire la qualità delle informazioni e la protezione delle persone. In questo ambito, particolare rilevanza assumono i *Data Labs*, concepiti come ambienti sicuri nei quali è possibile effettuare operazioni di preparazione, pulizia, validazione e pseudonimizzazione dei dati, al fine di aumentare l'affidabilità dei sistemi di intelligenza artificiale e ridurre il rischio di discriminazioni o *bias* sistemici. La loro funzione, centrale, nel modello europeo, si collega direttamente al principio di neutralità tecnologica, richiamato dalla dottrina costituzionale secondo cui la struttura del linguaggio algoritmico incide sulla conformità delle decisioni ai principi di eguaglianza, imparzialità e prevedibilità [28].

All'interno del *Digital Omnibus* emerge con chiarezza la volontà della Commissione di sostenere lo sviluppo tecnologico riducendo gli oneri amministrativi che gravano in particolare sulle piccole e medie imprese e sulle *small mid caps*. La semplificazione è intesa come strumento per promuovere un'applicazione proporzionata degli obblighi previsti dall'*AI Act*, senza alterare il livello delle tutele. Il testo propone l'ampliamento delle misure di alleggerimento documentale, l'introduzione di linee guida tecnico-contrattuali e il potenziamento dei meccanismi di sperimentazione regolatoria, tra i quali assumono particolare rilievo le *regulatory sandboxes*, concepite come contesti controllati nei quali

tenzialmente lesive dei diritti fondamentali e, al contempo, valorizzandone le potenzialità al servizio dell'economia e della società. Cfr. inoltre: G. FINOCCHIARO, *op. loc. cit.*; O. POLLICINO-P. DUNN, *op. loc. cit.*; A. PAJNO-F. DONATI-A. PERRUCCI (a cura di), *op. loc. cit.*; A. D'ALOIA, *Intelligenza Artificiale e diritto. Come regolare un mondo nuovo*, Milano, 2021.

[28] Sul nesso tra linguaggio algoritmico, neutralità tecnologica e principi di eguaglianza, imparzialità e prevedibilità v. D.-U. GALETTA-B. CAPPIELLO, *Il Regolamento UE sull'intelligenza artificiale e le amministrazioni pubbliche*, in A.F. URICCHIO-C. CALDAROLA (a cura di), *op. cit.*, p. 59 ss., con riferimento ai rischi di opacità decisionale e alla necessità di trasparenza.

testare soluzioni innovative garantendo, al contempo, la salvaguardia dei diritti degli utenti e la supervisione delle autorità competenti [29]. Tali strumenti, già impiegati in altri settori regolati, consentono un approccio flessibile alla disciplina, preservando il nucleo essenziale delle garanzie senza ostacolare l'evoluzione tecnologica [30].

L'equilibrio necessario tra innovazione e protezione dei diritti si manifesta inoltre nella struttura dell'*AI Act*, che impone requisiti di trasparenza, supervisione umana e tracciabilità dei processi decisionali per i sistemi ad alto rischio. Il regolamento istituisce l'*AI Office*, organismo sovranazionale dotato di funzioni di coordinamento, vigilanza e indirizzo interpretativo, destinato a garantire un'applicazione uniforme del quadro regolatorio e a supportare gli Stati membri nell'attuazione delle disposizioni relative ai modelli avanzati di intelligenza artificiale [31]. La scelta di una supervisione centralizzata risponde all'esigenza di evitare divergenze applicative che potrebbero compromettere l'unità del mercato e generare incertezze per imprese e cittadini, dovendo l'algoritmo operare all'interno di un perimetro giuridico ben delimitato [32].

La Commissione ammette altresì che la complessità del quadro regolatorio possa ostacolare l'innovazione, specialmente nei settori nei quali si cumulano obblighi di conformità derivanti da normative diverse. Per tale ragione il *Digital Omnibus* propone l'istituzione di un punto unico europeo di notifica degli inci-

[29] European Commission, *Proposal for a Regulation on Artificial Intelligence, Cybersecurity and Data Governance (Digital Omnibus)*, COM (2025) 837 final; sulla funzione degli strumenti di flessibilità e sperimentazione regolatoria nell'attuazione dell'*AI Act* v. C. CALDAROLA, *op. cit.*, p. 87 ss.

[30] M. EBERS, *Truly risk-based regulation of artificial intelligence: how to implement the EU's AI Act*, in *EJRR*, 16, 2025, pp. 684-703. L'A. osserva che l'emersione di approcci regolatori "risk-based" in materia di sistemi di intelligenza artificiale, tanto nell'Unione europea quanto in altri ordinamenti, solleva innanzitutto l'interrogativo circa il significato attribuito al concetto di «rischio» da parte del legislatore. In termini generali, il rischio può essere definito come «la probabilità che una fonte di pericolo si traduca in un danno effettivo». Si vedano anche C. NOVELLI-F. CASOLARI-A. ROTOLO-M. TADDEO-L. FLORIDI, *AI risk assessment: a scenario-based, proportional methodology for the AI Act*, in *Dig. Soc.*, 3:2, 2024; M. KAMINSKI, *Regulating the risks of AI*, in *Boston ULR*, 2023, p. 103.

[31] Reg. (UE) 2024/1689 (*AI Act*); sull'istituzione dell'*AI Office* e sul ruolo della supervisione centralizzata ai fini dell'applicazione uniforme del quadro regolatorio v. C. CALDAROLA, *op. cit.*, p. 91 ss.; M. CORRADINO, *op. cit.*, p. 39 ss.

[32] Sul principio della centralità della persona umana e sull'impossibilità di considerare l'algoritmo come soggetto autoreferenziale v. A.F. URICCHIO-F. MAZZOTTA, *Metaverso, fiscalità e intelligenza artificiale: prime considerazioni alla luce del recente disegno di legge in materia*, in A.F. URICCHIO-C. CALDAROLA (a cura di), *op. cit.*, p. 225 ss., ove il controllo umano è ricondotto a presidio essenziale della legalità costituzionale.

denti di cybersicurezza, destinato a sostituire la pluralità degli obblighi previsti dal GDPR, dalla direttiva NIS2 e dal Regolamento DORA, riducendo il carico amministrativo e migliorando la cooperazione tra le autorità nazionali [33]. Tale soluzione ha trovato riscontri positivi nelle prime analisi condotte in ambito tributario, dove si è osservato che la proliferazione degli adempimenti digitali richiede un equilibrio coerente tra efficienza operativa, certezza giuridica e garanzie procedimentali [34].

La visione europea tende dunque a identificare un modello di "innovazione governata", nel quale la tecnologia si sviluppa all'interno di una cornice normativa capace di assicurare certezza giuridica, sostenibilità e tutela dei diritti. La disciplina dell'intelligenza artificiale diviene l'ambiente giuridico nel quale confluiscono i diversi piani del diritto: quello pubblico, che fissa limiti e valori non negoziabili; quello privato, che disciplina responsabilità, rapporti contrattuali e circolazione dei beni digitali; quello della sicurezza, che protegge gli interessi essenziali dello Stato; e quello geopolitico, che inserisce l'intelligenza artificiale nel confronto tra poteri globali, quello tributario che da un lato apprezza e utilizza i dati per contrastare l'evasione fiscale , dall'altro è chiamato ad apprezzare nuove fattispecie imponibile.

Ne discende un forte impegno dell'Unione europea nel perseguire un equilibrio tra competitività tecnologica e tutela della persona, quale condizione per garantire a privati cittadini, imprese e istituzioni fiducia negli strumenti digitali, utilizzo responsabile, e quindi sicurezza e qualità dei dati.

4. La strategia regolatoria europea come strumento necessario per assicurare l'identità digitale e contrastare l'opacità algoritmica

La strategia regolatoria dell'Unione europea in materia di dati, intelligenza artificiale e identità digitale si inscrive entro una traiettoria che oltrepassa la dimensione della mera regolazione settoriale delle tecnologie emergenti, configurandosi come un progetto organico di sovranità digitale [35]. In tale prospettiva, la normazione europea non svolge soltanto una funzione ordinatrice o correttiva

[33] European Commission, *SWD (2025) 836 final – Digital Simplification Agenda*; sul coordinamento degli obblighi di notifica in materia di cybersicurezza e sulla riduzione della frammentazione regolatoria v. F. GALLO, *Cyberspazio, diritti e tutele*, cit., p. 23 ss.

[34] Sul rapporto tra semplificazione degli adempimenti digitali, certezza giuridica ed efficienza operativa, con specifico riferimento all'ambito tributario, v. A.F. URICCHIO, *Il fisco dinanzi all'intelligenza artificiale e alla digital transformation*, cit., p. 201 ss.

[35] Commissione europea, *Data Union Strategy – Unlocking Data for AI*, cit.

dei mercati, ma diviene lo strumento attraverso il quale l'Unione riafferma la propria capacità di governo delle trasformazioni tecnologiche, preservando l'unità del mercato interno e la coerenza dell'azione pubblica in un contesto caratterizzato dalla crescente pervasività delle decisioni automatizzate [36].

Nel contesto globale, l'approccio europeo si distingue per una marcata vocazione istituzionale [37]. A fronte di modelli che privilegiano, da un lato, la flessibilità regolatoria e l'autoregolazione del mercato, come nell'esperienza statunitense, e, dall'altro, il controllo politico centralizzato delle infrastrutture digitali, come nel modello cinese, l'Unione europea persegue una via fondata su regole vincolanti, standard armonizzati e meccanismi di supervisione sovranazionale [38]. Anche rispetto all'impostazione britannica (che assume i principi di *safety, transparency, fairness, accountability, contestability*), orientata a una regolazione per principi e alla responsabilizzazione delle autorità settoriali, il modello europeo si caratterizza per l'ambizione di costruire un quadro normativo unitario, idoneo a garantire uniformità applicativa e certezza del diritto in un ecosistema tecnologico strutturalmente transnazionale [39]. In questo quadro, la governance dei dati assume una valenza eminentemente strategica [40]. Gli spazi europei dei dati sono concepiti non come meri ambienti di scambio informativo, ma come infrastrutture di interesse generale, destinate a sostenere l'innovazione, l'integrazione del mercato interno e la modernizzazione dell'azione pubblica [41]. L'interoperabilità delle infrastrutture, l'adozione di standard comuni e la qualità semantica delle informazioni costituiscono presupposti essenziali affinché la circolazione dei dati possa tradursi in valore economico e sociale senza generare asimmetrie informative, frammentazioni regolatorie o rischi sistemici per il funzionamento del mercato unico [42]. In tale prospettiva, la qualità del dato non rappresenta un obiettivo meramente tecnico, ma una condizione strutturale di affidabilità complessiva degli ecosistemi digitali europei.

---

[36] Regolamento (UE) 2024/1689 (*AI Act*), considerando iniziali; SWD (2025) 836 final, Introduzione e § 1.1; G. DE MINICO, *op. loc. cit.*, Parte II.

[37] F. GALLO, *Cyberspazio, diritti e tutele*, in A.F. URICCHIO-C. CALDAROLA, *op. cit.*, p. 15 ss.

[38] F. GALLO, *op. loc. ult. cit.*; N. MACCABIANI, *Data processing and algorithms: a constitutionally oriented AI?*, in *AI Law*, 2025, p. 1 ss.

[39] Commissione europea, *Digital Omnibus*, COM (2025) 837 final, Explanatory Memorandum.

[40] Commissione europea, COM (2025) 835 final, §§ 1 e 3.

[41] Commissione europea, COM (2025) 835 final, § 2; Commission Staff Working Document on Common European Data Spaces, SWD (2024) 21 final.

[42] Commissione europea, COM (2025) 835 final, § 3; SWD (2025) 836 final, § 1.1.2.

La strategia regolatoria europea affronta il tema dell'opacità algoritmica come una questione eminentemente istituzionale, e non soltanto come problema di tutela individuale [43]. L'assenza di trasparenza e tracciabilità nei processi decisionali automatizzati, in quanto fattore di disomogeneità applicativa, incide direttamente sulla fiducia nel mercato, sulla capacità delle autorità pubbliche di esercitare funzioni di controllo e sulla stessa unità dell'ordinamento europeo [44]. In tale prospettiva, l'opacità algoritmica assume la fisionomia di un rischio sistemico, suscettibile di compromettere la coerenza applicativa delle norme e di generare divergenze tra Stati membri, con effetti distorsivi sulla concorrenza e sulla certezza del diritto [45].

È in questo contesto che si colloca l'AI Act, il quale introduce un modello regolatorio multilivello fondato sulla gestione del rischio, sulla trasparenza dei sistemi e sulla responsabilità degli operatori [46]. Particolarmente significativa è l'istituzione dell'AI Office, chiamato a svolgere funzioni di coordinamento, vigilanza e indirizzo interpretativo a livello sovranazionale [47]. La centralizzazione di tali competenze risponde all'esigenza di evitare frammentazioni applicative e di garantire un'attuazione uniforme del regolamento, assicurando che l'intelligenza artificiale operi entro un perimetro giuridico chiaro, prevedibile e controllabile [48].

La funzione dell'AI Office non si esaurisce, tuttavia, nella dimensione interna dell'ordinamento europeo, ma si proietta sul piano internazionale, contribuendo a consolidare il ruolo dell'Unione quale potenza normativa nel governo globale dell'intelligenza artificiale [49]. Attraverso la definizione di standard tecnici, criteri di valutazione del rischio e modelli di governance multilivello, l'Unione esercita una capacità di influenza che si fonda sulla forza del mercato unico e sulla coerenza del proprio impianto giuridico, offrendo un paradigma regolatorio alternativo in un contesto di crescente competizione tecnologica e geopolitica [50].

---

[43] A. SIMONCINI, *op. loc. cit.*; G. DE MINICO, *op. loc. cit.*

[44] Regolamento (UE) 2024/1689 (*AI Act*), artt. 13 e 50; SWD (2025) 836 final, § 1.2.2.7.

[45] F. GALLO, *op. loc. cit.*; Commissione europea, COM (2025) 837 final, motivazioni sulla necessità di uniformità applicativa del mercato interno.

[46] Regolamento (UE) 2024/1689 (AI Act); SWD (2025) 836 final, § 3.

[47] Regolamento (UE) 2024/1689 (AI Act), artt. 64 ss.; Commissione europea, COM (2025) 837 final; Comunicato stampa Commissione UE, 19 novembre 2025.

[48] Commissione europea, COM (2025) 837 final, *Explanatory Memorandum;* SWD (2025) 836 final, § 3.2.

[49] Commissione europea, COM (2025) 835 final, § 3.3.

[50] F. GALLO, *op. loc. cit.*; N. MACCABIANI, *op. loc. cit*.

In tale cornice si inserisce anche la disciplina dell'identità digitale delle imprese, che trova nei Portafogli Aziendali Europei uno strumento volto a rafforzare la certezza giuridica e la fiducia nelle transazioni economiche transfrontaliere [51]. La creazione di un'infrastruttura interoperabile per la gestione delle attestazioni giuridicamente rilevanti consente di ridurre gli oneri amministrativi, prevenire le frodi e favorire l'integrazione del mercato interno, contribuendo alla costruzione di un ambiente digitale affidabile e competitivo [52]. L'identità digitale assume così una funzione sistemica, quale elemento di raccordo tra innovazione tecnologica, sicurezza delle relazioni giuridiche e integrazione economica europea [53].

Nel loro insieme, tali strumenti delineano una strategia regolatoria che non si limita a reagire all'innovazione, ma ne orienta lo sviluppo secondo una visione di lungo periodo, nella quale qualità delle infrastrutture digitali, trasparenza dei processi decisionali e coerenza delle regole costituiscono condizioni imprescindibili per la competitività e la sostenibilità dell'economia europea [54]. La regolazione si configura, pertanto, non come vincolo esterno, ma come fattore abilitante, capace di rafforzare la fiducia negli ecosistemi digitali e di consolidare il ruolo dell'Unione europea quale attore centrale nel governo delle trasformazioni tecnologiche globali [55].

5. Conclusioni

La ricostruzione del quadro europeo dell'innovazione digitale mette in evidenza come l'Unione abbia maturato una visione nella quale la regolazione non costituisce un contrappeso alla tecnologia né un limite esterno alla sua espansione, bensì una condizione intrinseca della sua legittimità e della sua sostenibilità di lungo periodo. Dalla Data Union Strategy al Digital Omnibus, dall'AI Act al Data Act, dal Data Governance Act fino alla proposta sui Portafogli

[51] Commissione europea, Proposta di regolamento sui *European Business Wallets*, COM (2025) 838 final.

[52] Commissione europea, COM (2025) 838 final, Explanatory Memorandum; SWD (2025) 837 final, §§ 6-8.

[53] Commissione europea, COM (2025) 838 final; Allegato tecnico ai *European Business Wallets*.

[54] A.F. URICCHIO-C. CALDAROLA (a cura di), *op. loc. cit.*; Commissione europea, COM (2025) 835 final, Conclusioni.

[55] F. GALLO, *op. cit*., p. 15 ss.; Commissione europea, COM (2025) 837 final e COM (2025) 838 final.

Aziendali Europei, emerge un paradigma comune, poiché la trasformazione digitale può produrre valore economico e sociale soltanto se costruita su infrastrutture di fiducia, qualità dei dati, trasparenza dei processi e centralità della persona umana [56]. In tale prospettiva, la regolazione assume una funzione sistemica e definisce il perimetro entro il quale l'innovazione può svilupparsi senza degenerazioni tecnicistiche che poi finiscono per dare luogo a rendite di posizione, ad asimmetrie informative e soprattutto a erosione delle garanzie costituzionali.

I documenti COM e SWD del 2025 confermano che l'obiettivo dell'Unione non va ricercato nella proliferazione normativa, bensì nella costruzione di un insieme coerente e interoperabile di strumenti giuridici. Il Digital Omnibus inaugura una logica di manutenzione evolutiva del *rulebook* digitale, orientata a verificarne l'efficacia, a razionalizzarne le ridondanze e a garantire uniformità applicativa. L'operazione non è meramente tecnica, poiché la semplificazione diventa strumento di giustizia regolatoria: essa permette agli operatori di accedere a un quadro più chiaro e proporzionato, riduce i costi di conformità e accresce l'affidabilità dell'azione pubblica. Il punto unico europeo di notifica degli incidenti di cybersicurezza ne è un esempio particolarmente significativo, giacché non attenua il livello di protezione degli interessati, ma ne rafforza l'effettività attraverso una gestione più coordinata ed efficiente delle segnalazioni.

L'innovazione, in questa prospettiva, non è un fenomeno neutro, poiché incide sulla fisiologia delle decisioni pubbliche, sui rapporti di forza tra privati, sulla misurabilità della responsabilità, sull'efficienza amministrativa, sulla sicurezza collettiva e persino sugli equilibri geopolitici. La sostenibilità della transizione digitale esige pertanto che il diritto europeo svolga una funzione di presidio costante, capace di prevenire marginalizzazioni, automatismi lesivi dei diritti e concentrazioni incontrollate di potere informazionale. Di conseguenza, la tutela della persona e la qualità delle istituzioni si configurano come criteri ordinatori dell'intera infrastruttura digitale dell'Unione.

Il processo di regolazione dell'ecosistema digitale europeo rappresenta, in ultima analisi, una prova di maturità istituzionale, poiché chiama l'Unione a dimostrare che la transizione tecnologica può essere governata senza rinunciare alla centralità della persona e alla razionalità del diritto. La capacità di integrare valori costituzionali, esigenze del mercato e dinamiche geopolitiche in un quadro nor-

[56] Sul paradigma unitario della regolazione europea dell'innovazione digitale, fondato su infrastrutture di fiducia, qualità dei dati, trasparenza dei processi e centralità della persona, v. A.F. URICCHIO, *Presentazione*, in A.F. URICCHIO-C. CALDAROLA (a cura di), *op. cit.*, p. 9 ss., ove la regolazione è qualificata come condizione intrinseca della legittimità e della sostenibilità della trasformazione tecnologica.

mativo coerente costituisce il banco di prova della sua autorevolezza nel nuovo ordine digitale globale. Da tale capacità dipenderanno non soltanto la competitività dell'Europa, ma anche la qualità democratica dei processi decisionali che accompagneranno la trasformazione algoritmica delle società contemporanee.

Il futuro dell'Europa digitale dipenderà, pertanto, dalla capacità delle istituzioni di mantenere coerenza, visione e responsabilità nel governo della transizione tecnologica. L'innovazione non è interpretata come una corsa priva di direzione, ma come un percorso attraverso il quale l'Unione europea intende affermare un modello di sviluppo digitale umano-centrico, sostenibile e pienamente conforme ai valori democratici che ne costituiscono il fondamento. Se tale direzione sarà mantenuta e rafforzata, l'ordinamento europeo potrà non soltanto governare la propria trasformazione tecnologica, ma anche offrire un paradigma internazionale nel quale progresso, equità e sicurezza coesistano in modo armonico.

AI LAW

INTERNATIONAL REVIEW
OF ARTIFICIAL INTELLIGENCE LAW

La nuova realtà della giustizia predittiva

Rachele Ranieri

Dottoranda in Diritto romano, teoria degli ordinamenti e diritto privato del mercato nell'Università degli Studi di Roma "La Sapienza"

Abstract

(EN): The essay addresses predictive justice as a paradigmatic manifestation of the entry of artificial intelligence into the very core of the judicial function, critically examining whether algorithmic tools, grounded in the probabilistic analysis of vast bodies of statutory and case-law data, may contribute to the making of judicial decisions without undermining the axiological, constitutional, and systemic foundations of judging. By reconstructing the promises associated with this phenomenon (greater efficiency, speed, predictability, uniformity, transparency, and the potential reduction of litigation) the analysis shows how the aspiration to a calculable decision is situated within a structural tension between the technological rationalization of law and the irreducible specificity of the individual case. From this perspective, judicial decision-making cannot be reduced to the mere serial application of precedents, since judging entails interpretive activity, the selection of legally relevant facts, the balancing of interests, the duty to give reasons, the assessment of reasonableness, and engagement with constitutional, European, and international principles. In this light, the replacement of the human judge by the machine judge proves problematic not only at the technical level, but above all in terms of democratic legitimacy and the protection of fundamental rights: algorithmic standardization risks producing a trivialization of judgment, the crystallization of precedents, decisional opacity, the de-responsibilization of the judge, the compression of impartiality, and the hollowing out of the properly human dimension of decision-making, which consists in discernment, experience, intuition, and sensitivity to the singularity of the dispute.

Particular attention is also devoted to the problems of bias embedded in data and models, to the difficulty of making the rationale of algorithmic outcomes intelligible, to the tension between trade secrecy and transparency, as well as to the issue of civil liability for damage caused by automated decisions. The analysis then extends to certain specific fields (criminal law, tax law, and family law) in which the possible applications of AI take different forms: while tax litigation and certain economic aspects of family law appear, on the one hand, more susceptible to logics of standardization and calculation, criminal law and, more generally, all areas characterized by a high degree of evaluative judgment reveal, on the other hand, with particular clarity the unsuitability of a fully automated form of justice.

The essay thus reaches a clear conclusion: the model of automated justice appears neither desirable nor compatible with a constitutionally oriented conception of the judicial function. The only normatively and axiologically sustainable use of artificial intelligence is an auxiliary and complementary one, within a paradigm of assisted justice, or human in the loop, in which the algorithm

operates as a tool supporting the activity of the judge, without ever replacing the judge's decisional role, argumentative responsibility, and guarantor function in the making of a just decision.

(SP): El ensayo aborda la justicia predictiva como manifestación paradigmática de la irrupción de la inteligencia artificial en el núcleo mismo de la función jurisdiccional, y examina críticamente si instrumentos algorítmicos, basados en el análisis probabilístico de grandes volúmenes de datos normativos y jurisprudenciales, pueden contribuir a la formación de la decisión judicial sin menoscabar los presupuestos axiológicos, constitucionales y sistemáticos del juzgar. A partir de la reconstrucción de las promesas que acompañan a este fenómeno (mayor eficiencia, rapidez, previsibilidad, uniformidad, transparencia y potencial reducción de la litigiosidad), el análisis pone de relieve que la aspiración a una decisión calculable se inserta en una tensión estructural entre la racionalización tecnológica del derecho y la irreductible singularidad del caso concreto. Desde esta perspectiva, la decisión jurisdiccional no puede quedar reducida a una mera aplicación seriada de precedentes, pues juzgar implica actividad interpretativa, selección de los hechos jurídicamente relevantes, ponderación de intereses, deber de motivación, control de razonabilidad y confrontación con principios constitucionales, europeos e internacionales. En este marco, la sustitución del juez humano por el juez-máquina se revela problemática no solo en el plano técnico, sino sobre todo desde la perspectiva de la legitimidad democrática y de la tutela de los derechos fundamentales: la estandarización algorítmica corre el riesgo de provocar una banalización del juicio, la cristalización de los precedentes, opacidad decisoria, desresponsabilización del juez, merma de la imparcialidad y vaciamiento de la dimensión propiamente humana de la decisión, integrada por discernimiento, experiencia, intuición y sensibilidad ante la singularidad de la controversia.
Se presta asimismo particular atención a los problemas derivados de los sesgos inscritos en los datos y en los modelos, a la dificultad de hacer inteligible la ratio del resultado algorítmico, a la tensión entre secreto empresarial y transparencia, así como a la cuestión de la responsabilidad civil por los daños ocasionados por decisiones automatizadas. El análisis se proyecta después sobre determinados ámbitos específicos (derecho penal, derecho tributario y derecho de familia) en los que las posibilidades de aplicación de la IA se presentan de manera diversa: si, por un lado, la litigación tributaria y ciertos aspectos patrimoniales del derecho de familia parecen más expuestos a lógicas de estandarización y cálculo, por otro, el derecho penal y, en términos más generales, todos aquellos sectores caracterizados por una elevada densidad valorativa ponen especialmente de manifiesto la inidoneidad de una justicia íntegramente automatizada.
El trabajo alcanza así una conclusión nítida: el modelo de justicia automatizada no resulta ni deseable ni compatible con una concepción de la jurisdicción orientada constitucionalmente. El único uso normativamente y axiológicamente sostenible de la inteligencia artificial es el de carácter auxiliar y complementario, dentro de un paradigma de justicia asistida, o *human in the loop*, en el que el algoritmo actúe como instrumento de apoyo a la actividad del juez, sin sustituir en ningún caso su función decisoria, su responsabilidad argumentativa ni su papel de garantía en la construcción de una decisión justa.

(FR): L'article aborde la justice prédictive comme une manifestation paradigmatique de l'irruption de l'intelligence artificielle au cœur même de la fonction juridictionnelle, et examine de manière critique si des instruments algorithmiques, fondés sur l'analyse probabiliste de vastes ensembles de données normatives et jurisprudentielles, peuvent contribuer à l'élaboration de la décision judiciaire sans compromettre les présupposés axiologiques, constitutionnels et systémiques de l'acte de juger. À partir de la reconstruction des promesses associées à ce phénomène (efficacité accrue, célérité, prévisibilité, uniformité, transparence et réduction potentielle du contentieux), l'analyse met en évidence que l'aspiration à une décision calculable s'inscrit dans une tension structurelle entre la rationalisation technologique du droit et l'irréductible singularité du cas con-

cret. Dans cette perspective, la décision juridictionnelle ne saurait être réduite à une simple application sérielle des précédents, dès lors que juger implique une activité interprétative, la sélection des faits juridiquement pertinents, la mise en balance des intérêts, l'obligation de motivation, le contrôle de la raisonnabilité et la confrontation avec des principes constitutionnels, européens et internationaux. Dans ce cadre, la substitution du juge humain par le juge-machine se révèle problématique non seulement sur le plan technique, mais surtout au regard de la légitimité démocratique et de la protection des droits fondamentaux : la standardisation algorithmique risque d'engendrer une banalisation du jugement, une cristallisation des précédents, une opacité décisionnelle, une déresponsabilisation du juge, une atteinte à l'impartialité et un dépouillement de la dimension proprement humaine de la décision, faite de discernement, d'expérience, d'intuition et de sensibilité à la singularité du litige.
Une attention particulière est également accordée aux problèmes résultant des biais inscrits dans les données et les modèles, à la difficulté de rendre intelligible la ratio du résultat algorithmique, à la tension entre secret des affaires et transparence, ainsi qu'à la question de la responsabilité civile pour les dommages causés par des décisions automatisées. L'analyse s'étend ensuite à certains domaines spécifiques (droit pénal, droit fiscal et droit de la famille) dans lesquels les potentialités d'application de l'IA se présentent de manière différenciée: si, d'une part, le contentieux fiscal et certains aspects patrimoniaux du droit de la famille paraissent davantage exposés à des logiques de standardisation et de calcul, d'autre part, le droit pénal et, plus généralement, tous les secteurs caractérisés par une forte densité évaluative mettent particulièrement en lumière l'inadéquation d'une justice entièrement automatisée.
La contribution parvient ainsi à une conclusion nette : le modèle de la justice automatisée n'apparaît ni souhaitable ni compatible avec une conception constitutionnellement orientée de la juridiction. Le seul usage de l'intelligence artificielle qui soit normativement et axiologiquement soutenable est un usage auxiliaire et complémentaire, dans le cadre d'un paradigme de justice assistée, ou *human in the loop*, dans lequel l'algorithme agit comme un instrument d'appui à l'activité du juge, sans jamais se substituer à sa fonction décisionnelle, à sa responsabilité argumentative ni à son rôle de garantie dans l'élaboration d'une décision juste.

(DE): Der Beitrag behandelt die prädiktive Justiz als paradigmatische Erscheinungsform des Eindringens künstlicher Intelligenz in den innersten Kern der Rechtsprechungsfunktion und geht der Frage in kritischer Weise nach, ob algorithmische Instrumente, die auf der probabilistischen Analyse umfangreicher normativer und rechtsprechungsbezogener Datenbestände beruhen, zur Bildung richterlicher Entscheidungen beitragen können, ohne die axiologischen, verfassungsrechtlichen und systemischen Grundlagen des Richtens zu beeinträchtigen. Ausgehend von einer Rekonstruktion der Versprechen, die dieses Phänomen begleiten (größere Effizienz, Schnelligkeit, Vorhersehbarkeit, Einheitlichkeit, Transparenz und ein potenzieller Rückgang der Prozessbelastung) zeigt die Untersuchung, dass das Streben nach einer berechenbaren Entscheidung in einer strukturellen Spannung zwischen der technologischen Rationalisierung des Rechts und der irreduziblen Besonderheit des Einzelfalls steht. Aus dieser Perspektive kann die richterliche Entscheidung nicht auf eine bloß serielle Anwendung von Präzedenzfällen reduziert werden, denn das Richten umfasst interpretative Tätigkeit, die Auswahl rechtlich relevanter Tatsachen, die Abwägung widerstreitender Interessen, die Begründungspflicht, die Kontrolle der Angemessenheit sowie die Auseinandersetzung mit verfassungsrechtlichen, europäischen und internationalen Grundsätzen. Vor diesem Hintergrund erweist sich die Ersetzung des menschlichen Richters durch den Maschinenrichter nicht nur auf technischer Ebene, sondern vor allem im Hinblick auf demokratische Legitimation und den Schutz der Grundrechte als problematisch: Die algorithmische Standardisierung birgt die Gefahr einer Banalisierung des Urteilens, einer Verfestigung der Präjudizien, einer Intransparenz der Entscheidung, einer Entverantwortlichung des Richters, einer Beeinträchtigung der Unparteilichkeit und einer Aushöhlung der genuin menschlichen Dimension

des Entscheidens, die sich aus Unterscheidungsvermögen, Erfahrung, Intuition und Sensibilität für die Singularität des Rechtsstreits zusammensetzt.
Besondere Aufmerksamkeit wird ferner den Problemen gewidmet, die sich aus in Daten und Modellen angelegten Verzerrungen ergeben, der Schwierigkeit, die *ratio* des algorithmischen Ergebnisses nachvollziehbar zu machen, der Spannung zwischen Geschäftsgeheimnis und Transparenz sowie der Frage der zivilrechtlichen Haftung für Schäden, die durch automatisierte Entscheidungen verursacht werden. Die Analyse erstreckt sich sodann auf einige spezifische Rechtsgebiete (Strafrecht, Steuerrecht und Familienrecht), in denen sich die Anwendungsmöglichkeiten der künstlichen Intelligenz in unterschiedlicher Weise darstellen: Während einerseits das Steuerstreitverfahren und bestimmte vermögensrechtliche Aspekte des Familienrechts stärker für Logiken der Standardisierung und Berechnung anfällig erscheinen, zeigen andererseits das Strafrecht und ganz allgemein all jene Bereiche, die durch eine hohe wertende Dichte gekennzeichnet sind, mit besonderer Deutlichkeit die Ungeeignetheit einer vollständig automatisierten Justiz.
Der Beitrag gelangt somit zu einem klaren Ergebnis: Das Modell einer automatisierten Justiz erscheint weder wünschenswert noch mit einem verfassungsorientierten Verständnis der Rechtsprechung vereinbar. Der einzig normativ und axiologisch tragfähige Einsatz künstlicher Intelligenz ist ein unterstützender und komplementärer, innerhalb eines Paradigmas assistierter Justiz oder des *human in the loop*, in dem der Algorithmus als Hilfsmittel für die Tätigkeit des Richters fungiert, ohne jemals dessen Entscheidungsrolle, argumentative Verantwortung und Garantiefunktion bei der Hervorbringung einer gerechten Entscheidung zu ersetzen.

(PT): O ensaio aborda a justiça preditiva como manifestação paradigmática da irrupção da inteligência artificial no próprio núcleo da função jurisdicional, examinando criticamente se instrumentos algorítmicos, baseados na análise probabilística de vastos conjuntos de dados normativos e jurisprudenciais, podem contribuir para a formação da decisão judicial sem comprometer os pressupostos axiológicos, constitucionais e sistémicos do julgar. Partindo da reconstrução das promessas que acompanham este fenómeno (maior eficiência, celeridade, previsibilidade, uniformidade, transparência e potencial redução do contencioso), a análise evidencia que a aspiração a uma decisão calculável se inscreve numa tensão estrutural entre a racionalização tecnológica do direito e a irredutível singularidade do caso concreto. Nesta perspetiva, a decisão jurisdicional não pode ser reduzida a uma mera aplicação seriada de precedentes, uma vez que julgar implica atividade interpretativa, seleção dos factos juridicamente relevantes, ponderação de interesses, dever de fundamentação, controlo da razoabilidade e confronto com princípios constitucionais, europeus e internacionais. Neste quadro, a substituição do juiz humano pelo juiz-máquina revela-se problemática não apenas no plano técnico, mas sobretudo à luz da legitimidade democrática e da tutela dos direitos fundamentais: a estandardização algorítmica corre o risco de provocar uma banalização do juízo, a cristalização dos precedentes, opacidade decisória, desresponsabilização do juiz, compressão da imparcialidade e esvaziamento da dimensão propriamente humana da decisão, feita de discernimento, experiência, intuição e sensibilidade à singularidade do litígio.
Concede-se, além disso, especial atenção aos problemas decorrentes dos enviesamentos inscritos nos dados e nos modelos, à dificuldade de tornar inteligível a ratio do resultado algorítmico, à tensão entre segredo comercial e transparência, bem como à questão da responsabilidade civil pelos danos causados por decisões automatizadas. A análise estende-se depois a alguns domínios específicos (direito penal, direito tributário e direito da família) nos quais as potencialidades de aplicação da IA se apresentam de forma diferenciada: se, por um lado, o contencioso tributário e alguns aspetos patrimoniais do direito da família parecem mais expostos a lógicas de estandardização e cálculo, por outro, o direito penal e, mais em geral, todos

os setores caracterizados por uma elevada densidade valorativa evidenciam com particular clareza a inadequação de uma justiça integralmente automatizada.
O trabalho chega, assim, a uma conclusão nítida: o modelo de justiça automatizada não se afigura nem desejável nem compatível com uma conceção constitucionalmente orientada da jurisdição. O único uso da inteligência artificial que se mostra normativamente e axiologicamente sustentável é o de carácter auxiliar e complementar, no quadro de um paradigma de justiça assistida, ou *human in the loop*, em que o algoritmo atua como instrumento de apoio à atividade do juiz, sem jamais substituir a sua função decisória, a sua responsabilidade argumentativa nem o seu papel de garantia na construção de uma decisão justa.

(IT): Il saggio affronta il tema della giustizia predittiva quale espressione paradigmatica dell'ingresso dell'intelligenza artificiale nel cuore della funzione giurisdizionale, interrogandosi criticamente sulla possibilità che strumenti algoritmici, fondati sull'analisi probabilistica di grandi masse di dati normativi e giurisprudenziali, possano concorrere alla formazione della decisione giudiziale senza compromettere i presupposti assiologici, costituzionali e ordinamentali del giudicare. Muovendo dalla ricostruzione delle promesse che accompagnano tale fenomeno (maggiore efficienza, rapidità, prevedibilità, uniformità, trasparenza e potenziale deflazione del contenzioso), l'indagine mostra come l'aspirazione a una decisione calcolabile si collochi entro una tensione strutturale tra razionalizzazione tecnologica del diritto e irriducibile specificità del caso concreto. A tal stregua la decisione giurisdizionale non può essere ridotta a mera applicazione seriale di precedenti, poiché il giudizio implica attività interpretativa, selezione dei fatti rilevanti, ponderazione degli interessi, motivazione, controllo di ragionevolezza e confronto con principi costituzionali, europei e internazionali. In questa prospettiva, la sostituzione del giudice-uomo con il giudice-macchina risulta problematica non soltanto sul piano tecnico, ma soprattutto sul piano della legittimazione democratica e della tutela dei diritti fondamentali: la standardizzazione algoritmica rischia di produrre banalizzazione del giudizio, cristallizzazione dei precedenti, opacità decisionale, deresponsabilizzazione del magistrato, compressione dell'imparzialità e svuotamento della dimensione propriamente umana del decidere, fatta di discernimento, esperienza, intuizione e sensibilità per la singolarità della controversia.
Particolare rilievo è attribuito, inoltre, ai problemi dei *bias* inscritti nei dati e nei modelli, alla difficoltà di rendere intelligibile la *ratio* dell'esito algoritmico, alla tensione tra segreto industriale e trasparenza, nonché alla questione della responsabilità civile per i danni prodotti da decisioni automatizzate. L'analisi si estende poi ad alcuni ambiti specifici (penale, tributario e familiare) nei quali le potenzialità applicative dell'IA si presentano in forme differenziate: se, da un lato, il contenzioso tributario e alcuni profili economici del diritto di famiglia sembrano più esposti a logiche di standardizzazione e calcolo, dall'altro lato il diritto penale e, in generale, tutte le aree ad alta densità valutativa mostrano con particolare evidenza l'inidoneità di una giustizia integralmente automatizzata.
Il contributo approda così a una conclusione netta: il modello della giustizia automatizzata non appare né auspicabile né compatibile con una concezione costituzionalmente orientata della giurisdizione. L'unico impiego normativamente e assiologicamente sostenibile dell'intelligenza artificiale è quello di tipo ausiliario e complementare, entro un paradigma di giustizia assistita o *human in the loop*, nel quale l'algoritmo operi come strumento di supporto all'attività del giudice, senza mai sostituirne il ruolo decisorio, la responsabilità argomentativa e la funzione di garanzia nella costruzione di una decisione giusta.

Sommario: 1. Premessa. – 2. La nuova era della giustizia predittiva. – 3. I benefici. – 4. I rischi connessi all'introduzione della giustizia predittiva. – 5. Il diritto penale. – 6. Il diritto tributario. – 7. Il diritto di famiglia – 8. Al di là del bene e del male. – 9. Conclusione.

1. Premessa

Amore, passione, tentazione, sofferenza e morte. Aspetti, tutti, che esprimono quel particolare e ambiguo rapporto che intercorre, ad oggi, tra il mondo del diritto e l'intelligenza artificiale. La seduzione della nuova scoperta e del progresso, in una società in continuo divenire, si è spinta a tal punto da suscitare l'idea, prima inimmaginabile neanche nei sogni più reconditi, di un mondo in cui la figura di «giudice robot» si sostituisca a quella di «giudice uomo». Ma un sistema c.d. di giustizia predittiva, se in apparenza sembrerebbe «accarezzare» l'apparato giurisdizionale con una mano sicura e di protezione, allo stesso tempo, tuttavia, cela nel profondo gravosi rischi e pericoli che potrebbero sfociare in una definitiva perdita della Giustizia sostanziale. La sfida di oggi e di domani è quella di creare un sistema in cui la macchina e l'uomo traggano beneficio l'un l'altro e lavorino insieme per un ordinamento migliore.

2. La nuova era della giustizia predittiva

In un mondo e in una società che sono in continua evoluzione dal punto di vista tecnologico, il campo del diritto, così come quello più specifico del processo, sono inevitabilmente costretti a confrontarsi con le potenzialità della rivoluzione e con i rischi associati all'uso dell'intelligenza artificiale [1]. Come premessa (che, in realtà, potrebbe suonare come un'eccessiva diffidenza nei confronti del progresso), ciò che alcuni filosofi avevano predetto potrebbe essere

[1] Com'è stato efficacemente osservato, «i robot entreranno a far parte prepotentemente della nostra vita quotidiana. La loro interazione con noi sarà profonda poiché, al contrario della rivoluzione informatica (che ne ha costituito una premessa tecnologica e culturale fondamentale) la robotica è fisicamente pervasiva e implica un sistematico contatto (in senso letterale) tra uomo e macchina. I robot guideranno auto, aerei e navi. Saranno presenti nelle nostre case e interagiranno con i nostri cari (in particolare anziani e bambini) gestiranno in modo autonomo interi settori della produzione industriale, affiancheranno (o forse sostituiranno!) i chirurghi umani al tavolo operatorio» (A. CARCATERRA, *Machinae autonome e decisione robotica*, in A. CARLEO (a cura di), *Decisione robotica*, Bologna, 2019, p. 50).

considerato come avveratosi, ovvero che «la tecnologia, intesa come universo delle tecnologie, e la razionalità che presiede alla loro funzionalità ed efficienza, non è più un mezzo a disposizione dell'uomo, ma è l'ambiente in cui anche l'Uomo subisce modifiche [2]. Avvicinandosi allo studio di tali fenomeni, è innegabile che fornire una definizione del termine «Intelligenza Artificiale» [3] (IA), tanto in termini legali quanto tecnici, sia un'operazione complessa. Recentemente, è stato proprio il legislatore europeo, con il Regolamento n. 1689/2024, il cosiddetto *Artificial Intelligence Act* (*AI Act* o *AIA*), uno tra i primi a fornire una definizione normativa.

Per ciò che qui interessa, di particolare attenzione appaiono le affermazioni di una nota giurista e, per tale ragione, si ritiene di poter approcciare il nostro tema mediante un richiamo ad esse. Nello specifico, l'Autrice ha sostenuto che «Nel tempo della Rivoluzione digitale, tra le molteplici promesse della tecnica, ve n'è una che risulta particolarmente seducente per il giurista: ripristinare la certezza del diritto dispersa nella complessità senza verità che caratterizzerebbe la contemporaneità. Grazie alla strabiliante evoluzione dell'IA, la c.d. giustizia algoritmica sarebbe prevedibile e, perciò, capace di superare la grande aporia iscritta nel doppio volto della società contemporanea: società liquida e dell'incertezza, da un lato; della calcolabilità universale, dall'altro» [4].

È, tuttavia, innegabile che è proprio nel momento giurisdizionale che le tensioni tra innovazione tecnologica e garanzie costituzionali raggiungono il loro apice dialettico.

Parlare di intelligenza artificiale nel settore della giustizia implica riferirsi a una molteplicità indeterminata, e forse ancora inimmaginabile, di macchine, strumenti e *software* potenzialmente utilizzabili [5]. Nel processo di modernizzazione della società, l'intelligenza artificiale ha svolto un ruolo primario per

[2] Argomentazione formulata da A. GALIMBERTI, *Psiche e techne. L'uomo nell'età della tecnica*, Milano, 2016.

[3] La paternità del termine è assegnata al matematico John McCarthy, il quale nel 1956 ha definito l'*artificial intelligence* come «la scienza e l'ingegneria della costruzione di macchine intelligenti». Tuttavia, molti considerano A. Turing il padre dei computer intelligenti e di ciò che in seguito sarebbe diventata l'intelligenza artificiale.

[4] R. MESSINETTI, *Brevi note sulla certezza del diritto nella società algoritmica*, in *Contr. impr.*, 2024, p. 283.

[5] Il *focus*, già da tempo avvertito nell'ampio e variegato dibattito sulle nuove e moderne tecnologie riguardanti il settore giuridico, è stato oggetto di riflessione da parte di molteplici autori. Tra i molti, si pensi a G. SARTOR, *L'intelligenza artificiale e il diritto. Un'introduzione*, Milano, 1996. Tra i più recenti E. BASSOLI, *Algoritmica giuridica. Intelligenza artificiale e diritto*, Milano, 2022.

anni, come evidente nella digitalizzazione dei servizi, nella creazione di *database*, nei sistemi che consentono l'identificazione precoce di casi ad alto rischio o prioritari – consentendo agli uffici giudiziari di allocare le risorse in modo più efficiente – e nella digitalizzazione dei ricorsi. L'intelligenza artificiale è quindi ufficialmente entrata in quello che Satta ha chiamato il «mistero del processo» [6].

Vogliamo ora richiamare l'attenzione su un'ipotesi specifica di applicazione dell'intelligenza artificiale nel campo del diritto, ovvero la trasformazione della fase del processo in un meccanismo di giustizia puramente predittiva. Come la miglior dottrina non ha mancato di sottolineare, «Ciò che è qui in gioco è l'immagine stessa della giustizia, ovvero, in altri termini, l'accettabilità sociale del diritto e della decisione giudiziaria che lo attua sul piano della realtà vitale. Ne consegue che il tema evocato dalla decisione algoritmica è il riconoscimento sociale del diritto, la sua appropriazione collettiva come parte della produzione culturale della società. Esso, perciò, coinvolge la fiducia, condizione necessaria del legame sociale e della tenuta di un dato ordine della vita associata» [7].

Un sistema algoritmico di tal specie, nello specifico, promuove la possibilità di prevedere gli esiti procedurali, ovvero la decisione giudiziaria, attraverso un calcolo probabilistico in cui il metodo scientifico viene applicato al diritto. Nel campo della prevedibilità, l'intelligenza artificiale è davvero una risorsa preziosa, in quanto può analizzare rapidamente e in modo efficiente enormi quantità di dati per derivare modelli e tendenze che possono aiutare a prevedere gli esiti dei tribunali. A ben vedere, questo fenomeno è reso possibile dalla capacità di «manipolazione» tecnologica con una velocità umanamente inaccessibile di enormi quantità di *big data*, estremamente eterogenei e non archiviabili dagli esseri umani.

Anche sul fronte dei valori correlati al *modus operandi* dei giudici, le mutazioni *in itinere* sono evidenti e di notevole impatto. È lecito preconizzare come stia radicalmente (e velocemente) mutando la stessa percezione connessa alla logica interpretativa delle norme e dei fatti giuridici. Come parte della dottrina ha autorevolmente affermato, «Tali principi possono essere trapiantati avendo riguardo all'azione del giudice che non è scevra dal contesto sociale attuale,

[6] S. SATTA, *Il mistero del processo*, in *Riv. dir. proc.*, 1949, p. 273 ss. (testo della conferenza tenuta all'Università di Catania il 4 aprile 1949, poi ripubblicato in ID., *Soliloqui e colloqui di un giurista*, Padova, 1968, pp. 3-18, sul quale v., in particolare, A. DE CUPIS, *Leggendo i 'Soliloqui e colloqui' di Salvatore Satta*, in *Riv. dir. civ*., 1969, II, p. 534 ss.); e in ID., *Il mistero del processo*, Milano, 1994, pp. 11-37).

[7] Argomentazione proposta da R. MESSINETTI, *op. cit*., p. 284.

contesto che viene, poi, ulteriormente complicato dall'insufficienza delle tradizionali categorie concettuali del diritto eguale, generale ed astratto, "un tempo baluardo dell'identità e indipendenza del giurista"» [8].

Ciò che appariva un mero ideale dell'Illuminismo ora sembra essere diventata una «realtà storica» [9]. Tuttavia, la questione, prima ancora di essere se ciò sia praticamente fattibile, è se sia auspicabile. Infatti, è innegabile che la giustizia predittiva solleva questioni su almeno due livelli. In primo luogo, in termini di legittimità e legalità e, in secondo luogo, in termini di qualità e affidabilità. In particolare, lo sviluppo dell'uso di sistemi di questo tipo presuppone l'accettazione del fatto che un algoritmo di apprendimento automatico – mediante tecniche di elaborazione del linguaggio naturale e apprendimento profondo – analizzi un numero indefinibile a priori di *big data* (rappresentati, ad esempio, da disposizioni normative e precedenti giurisprudenziali). Questo al fine di valutare, sulla base di un calcolo statistico e di un ragionamento induttivo-probabilistico, la validità o l'infondatezza di una determinata soluzione giuridica in relazione a una specifica questione di fatto o di diritto controverso. Si vorrebbe, quindi, sostituire il giudice-uomo con il giudice-robot.

3. I benefici

Prima di procedere con l'analisi di quelli che, secondo la maggioranza degli studiosi, sono i motivi che ostacolano - non solo con riguardo ai nostri giorni, ma in modo definitivo - l'adozione di un sistema di giustizia predittiva puro, è opportuno prendere atto di quelli che, allo stesso tempo, sono i vantaggi che il sistema giuridico potrebbe trarre (e che timidamente trae oggi) dall'uso dell'Intelligenza Artificiale in funzione decisoria [10].

---

[8] C. ROMEO, *Processo del lavoro e i nuovi stilemi tra legge e giurisprudenza*, in *Lav. giur.*, 2026, p. 5. L'A. a sua volta richiama nelle note P. BARCELLONA, *L'uso alternativo del diritto*, I, *Scienza giuridica e analisi marxista*, e II, *Ortodossia giuridica e pratica politica*, Roma-Bari, 1973, il quale riteneva che «la lettura dell'esperienza conduce inesorabilmente a verificare la crisi della mediazione giuridica, ormai divenuta anche crisi del giurista alla ricerca della sua perduta identità».

[9] B. OCCHIUZZI, *Algoritmi predittivi: alcune premesse metodologiche*, in *Dir. pen. cont.*, 2019, II, p. 393.

[10] Se ne trova rassegna in AA.VV., *Tavola rotonda. La c.d. Giustizia Predittiva: esperienze e prospettive*, in U. SALANITRO (a cura di), *SMART la persona e l'infosfera.* Atti del Convegno 30 settembre - 2 ottobre 2021 Catania, Pisa, 2022, p. 431 ss.

La predittività applicata alla giustizia non è un tema nuovo, come, infatti, evidenziano alcuni scritti di Poisson [11] e Condorcet [12]. In tempi più recenti, la questione è stata affrontata nei suoi aspetti più diversi nella maggior parte dei sistemi giuridici nazionali. Si pensi, a titolo di esempio, agli studi di Loevinger sulla cosiddetta giurisprudenza intelligente [13] e all'indagine di Reed C. Lawlor sull'utilità dell'uso di computer e algoritmi nella ricerca legale scientifica e nei processi [14].

Oppure ancora alle conferenze internazionali sull'operatività dei cosiddetti sistemi esperti [15], tenutesi rispettivamente a Bologna e Firenze, e, più recentemente, all'introduzione in Francia della legge n. 1321/2016 (la cosiddetta «Legge per la Repubblica digitale») e all'istituzione della piattaforma online *Prédictice* [16]. Anche in Italia si registrano i primi esempi di sperimentazione dell'IA nel settore della giustizia, come Prodigit, un progetto nel campo fiscale che mira a creare sistemi in grado di analizzare leggi, sentenze e contributi dottrinali per prevedere decisioni giuridiche. Di non minore importanza è l'iniziativa promossa nel 2021 dal Tribunale e dall'Università di Brescia per fornire agli utenti tutti gli elementi del «Diritto del lavoro» e del «Diritto societario», al fine di scoraggiare cause temerarie e promuovere la trasparenza delle decisioni. O l'accordo di collaborazione tra il CED della Corte di Cassazione e la Scuola Superiore Universitaria di Pavia (IUSS).

In effetti, è innegabile che un sistema capace di sfruttare i frutti dell'evoluzione tecnologica guadagna (sebbene con le precisazioni che seguiranno) in termini di certezza, efficacia, efficienza, trasparenza, velocità e deflazione del contenzioso. Si pensi alla circostanza per la quale i professionisti del diritto potrebbero avere un accesso più ampio e, soprattutto, più rapido alle informazioni. Sarebbe loro di grande ausilio una maggiore conoscenza e comprensione della situazione normativa, giurisprudenziale e fattuale inerente al caso.

---

[11] S.D. POISSON, *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*, Paris, 1837.

[12] M. DE CONDORCET, *Essai sur application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Paris, 1785.

[13] L. EVINGER, *Jurimetrics: The Next Step Forward*, in *Minnesota LR*, 1949, p. 455 ss.

[14] C. LAWLOR, *What Computers Can Do: Analysis and Prediction of Judicial Decisions*, in *ABA J*, 1963, p. 337 ss.

[15] L. LOMBARDI VALLAURI, *Verso un sistema esperto giuridico integrale*, in *Jus*, 1996, p. 157 ss.

[16] Lo strumento viene utilizzato per calcolare la probabilità di una causa di essere risolta e l'importo dei danni ottenuti in cause simili, su una banca dati costituita dalle sentenze delle Corti d'appello e della Corte Suprema (www.predictice.com/fr).

La società è affascinata dall'idea che l'intelligenza artificiale possa fornire risposte scevre da errori, che potrebbero, invece, condizionare la valutazione dell'essere umano e raggiungere un esito giudiziario più affidabile perché privo di influenze morali, etiche, valoriali e politiche. In forza di un presunto principio di non discriminazione algoritmica, il titolare del trattamento, ovvero, in questo caso, il giudice, dovrebbe utilizzare «procedure matematiche o statistiche appropriate per la profilazione» [17] attraverso la predisposizione di misure volte a impedire il verificarsi di effetti discriminatori nei confronti delle persone fisiche sulla base della razza o dell'origine etnica, delle opinioni politiche, della religione o delle convinzioni personali, dell'appartenenza sindacale, dello *status* genetico, dello stato di salute o dell'orientamento sessuale [18].

In caso di controversie ovvero di aspetti procedurali o sostanziali individuali, un sistema intelligente aumenterebbe certamente la calcolabilità o, in altre parole, la prevedibilità delle decisioni. Ciò, in particolare, rappresenta una declinazione più specifica del principio di certezza del diritto. È, infatti, innegabile, che, ad oggi, il diritto privato moderno, alla stregua delle sue peculiari esigenze ermeneutiche, non abbia mai perseguito la certezza in termini assimilabili alla calcolabilità matematica [19]. Dalle parole di Portalis sembra che ciò sia conseguenza del fatto che il calcolo matematico non apparirebbe compatibile con la natura del diritto, dell'interpretazione e della giustizia [20].

Non sarebbe meno vantaggioso in termini di tempo, poiché l'uso delle macchine ridurrebbe notevolmente la «durata dei procedimenti» in modo da garantire una velocità di giudizio totalmente nuova (e sperata dalla maggior parte delle persone). Allo stesso tempo, l'intelligenza artificiale potrebbe suggerire agli avvocati e ai cittadini di evitare le soluzioni giudiziarie a favore di nuove opportunità di risoluzione delle controversie fuori dai Tribunali, come la nego-

[17] Così Cons. Stato, sez. IV, 4 febbraio 2020, n. 881.

[18] Il principio di eguaglianza si traduce nel riconoscere e nell'eliminare qualsiasi elemento distorto o sbagliato che possa essere contenuto nei dati raccolti senza che tale operazione di bonifica delle informazioni comprometta l'efficienza e la precisione del sistema di AI. La giurisprudenza, specialmente quella in materia di giuslavoristica, con riguardo ai *riders* ha ravvisato una forma di discriminazione algoritmica dell'algoritmo Frank che pone sullo stesso piano i lavoratori assenti per motivi giustificati con quelli assenti, invece, per motivi non giustificati: così l'ordinanza Trib. Bologna, 31 dicembre 2020. Si vedano i commenti di M. LAMMANNIS, *Frank è un falso cieco: l'algoritmo discrimina i riders. Nota a Tribunale di Bologna, Sez. lav., 31 dicembre 2020*, *ord.*, in *Lav. giur*., 2021, p. 5; G. PEZZINI, *L'algoritmo Frank è cieco, ma il datore di lavoro ci vede benissimo*, in *EQUAL*, 9 gennaio 2022.

[19] Senza alcuna pretesa di esaustività, si possono richiamare N. BOBBIO, *Il positivismo giuridico*, Torino, 1979, p. 84 ss.; R. MESSINETTI, *op. cit*., p. 286.

[20] È il *Discours préliminaire du premier project de Code Civil* (21 gennaio 1801).

ziazione o la mediazione. Inoltre, potrebbero essere istituiti (o, come già avviene in alcune legislazioni nazionali, aumentati) sistemi e meccanismi per il calcolo degli indennizzi in materia civile (ad esempio, indennizzo per lesioni personali, indennizzo per licenziamento). Ultimo ma non meno importante vantaggio sarebbe che le decisioni si baserebbero su un esame più «oggettivo, neutrale, completo e accurato dei documenti e delle informazioni in essi contenute e sarebbero anche meno esposte ai rischi di «rumore, ovvero soggettivismo giudiziario» [21].

Pare, dunque, che attraverso la tecnologia si possa recuperare quella certezza e calcolabilità nella gestione dei rapporti, pretesa dal mercato, che il diritto non riesce più ad assicurare sufficientemente, anestetizzando la discrezionalità ermeneutica del giudice.

Sembra appropriato soffermarsi per un momento sul beneficio che il sistema trarrebbe in termini di trasparenza dalla diffusione pubblica di sentenze e ordinanze giudiziarie sui portali web delle autorità giudiziarie. Sebbene ciò, in effetti, svolga un ruolo fondamentale nel garantire la libera ricerca di tutte le decisioni giudiziali, la loro piena divulgazione espone, tuttavia, le persone fisiche coinvolte a rischi individuali di violazione dei diritti, tra i quali emerge il diritto alla privacy, e delle libertà fondamentali. Parte della dottrina ha tentato di ovviare a questo rischio e, quindi, di conciliare il necessario rispetto per i diritti degli individui con le esigenze della circolazione delle informazioni in uno Stato democratico. La soluzione si può trovare nel cosiddetto diritto all'anonimizzazione delle sentenze [22], che implica la rimozione, d'ufficio o su richiesta di una delle parti, dei dati personali contenuti nel testo di una sentenza o di un'ordinanza del Tribunale prima che venga resa pubblica.

Una volta «ripuliti» da qualsiasi riferimento personale, i documenti potrebbero essere inseriti in un sistema di giustizia predittiva che fornisce una valutazione della probabilità di vincere una controversia o di dedurre automaticamente massime di diritto.

4. I rischi connessi all'introduzione della giustizia predittiva

Pur consci dei numerosi e innegabili vantaggi che il sistema giudiziario potrebbe trarre dall'uso dei sistemi di giustizia predittiva, è necessario evidenziare

[21] Si v. sul tema D. KAHNEMAN-O. SIBONY-R. CASS SUNSTEIN, *Rumore*, Torino, 2021, p. 23.

[22] Il diritto all'anonimizzazione delle sentenze è disciplinato dal Regolamento (UE) 2016/679 e dall'articolo 52 del Codice della privacy.

adeguatamente quelli che, allo stesso tempo, e al momento certamente di maggiore peso e rilevanza, sono i rischi, i limiti e le illusioni connessi alla sostituzione della figura umana del giudice con una macchina. Come vedremo in dettaglio, gli ostacoli e i pericoli sono così tanti che superano i benefici. Prima di procedere, si può riassumere questo pensiero evocando l'emblematica espressione coniata da G. Calabresi per descrivere il progresso. L'Autore, in particolare, parla di questo riferendosi a un «dono dello spirito maligno», che si offre all'umanità per rendere la sua vita sempre più confortevole, ma che nel tempo si rivela tutt'altro che gratuito, imponendo costi elevati la cui correlazione con i benefici non è immediatamente percepibile [23].

Come si avrà modo di vedere, andando ad approcciare a una nuova realtà di una giustizia intelligente ci si scontra con un quadro complesso e sfaccettato, in cui l'entusiasmo di fronte all'innovazione tecnologica risulta sensibilmente temperato dal timore che l'intelligenza artificiale possa, in un certo senso, stravolgere il modo di guardare al diritto e, specificamente, alla funzione giurisdizionale.

Da questa prospettiva, poiché la giustizia predittiva è uno strumento che mira a sconfiggere e spodestare il giudice in carne e ossa, la stessa deve necessariamente essere esaminata analizzando la sua compatibilità con il quadro giuridico di derivazione internazionale, europea e nazionale. In effetti, un uso improprio di questa tecnologia potrebbe rappresentare un pericolo non solo per i diritti e le libertà umane, ma anche per l'ordine democratico in generale.

Un primo limite insormontabile per un trasferimento totalizzante del potere decisionale dall'uomo alla macchina si trova nella specificità dei singoli casi. In altre parole, per usare un'espressione di N. Irti, non esiste ripetitività nel diritto [24]. Ogni caso della vita quotidiana differisce da un altro, perché i soggetti, le cause, i sentimenti, le passioni e altri aspetti infiniti che hanno contribuito all'emergere di quel rapporto o evento sono tra loro diversi [25].

---

[23] G. CALABRESI, *Il dono dello spirito maligno: gli ideali, le convinzioni, i modi di pensare nei loro rapporti col diritto*, Milano, 1996. Proprio con una particolare variante di questo «dono» si concludono le pagine di F. ADDIS, Laudatio *di Guido Calabresi*, in *Pers. merc.*, 2014, p. 212.

[24] Per una più ampia trattazione si rimanda a N. IRTI, *Un diritto incalcolabile*, Torino, 2015.

[25] Ad avviso di A. SIMONCINI, *La dimensione costituzionale della giustizia predittiva. Riflessioni su intelligenza artificiale e processo*, in *Riv. dir. proc.*, 2024, p. 389, «nel mondo della giustizia predittiva, dominato da algoritmi di tipo statistico, non è possibile un mutamento di giurisprudenza, perché l'insieme delle decisioni già prese – sulla cui base è stato addestrato l'algoritmo decisionale – finisce per vincolare in modo assoluto e necessitato le nuove decisioni».

Da ciò consegue che, se la macchina trae la sua decisione dall'analisi delle decisioni passate su casi simili, non sarà mai in grado di trovare una risposta, poiché non ha e non esiste un «caso (completamente) analogo» da cui tradurla.

Strettamente correlato al limite appena menzionato, come ulteriore e fondamentale ostacolo alla «meccanizzazione» del processo, è opportuno soffermarsi sul ruolo del giudice. In effetti, l'attività da lui svolta non si limita mai a un'operazione puramente meccanica di applicazione della legge [26]. E, ciò nonostante, la vigenza del tradizionale principio giuspositivistico del giudice come «bouche de la loi» [27], il quale, tuttavia, ha ormai acquisito una valenza più simbolica che autenticamente pratica [28]. Si pensi, ad esempio, a tutte quelle decisioni, anche puramente algoritmiche, che, pur essendo conformi all'applicazione del principio astratto, vengono, per certe ragioni, istintivamente rifiutate dall'uomo medio poiché, per esempio, contrari a quello che nel mondo anglosassone viene definito «common sense».

Tuttavia, è necessario fare attenzione a non incorrere in un malinteso. In realtà, quanto appena detto non è volto ad attribuire al decisore una funzione legislativa che, secondo la tripartizione di Montesquieu, spetta al Parlamento.

Ma, al massimo, si intende rendere le persone consapevoli di una realtà indiscutibile, ovvero che la Giurisprudenza, anche se non la crea, «fa la legge» [29]. In primo luogo, si pensi alle società moderne, le quali sono in costante evoluzione e le cui dinamiche sono spesso troppo fluide e vivaci per essere previste e adeguatamente regolate da regole predeterminate e specifiche, anche a causa del lungo tempo tecnico necessario per approvare una legge. Si parla, quindi, di

[26] Pone particolare attenzione sul termine «meccanica» E. TAGLIASACCHI, *Il sindacato del giudice sull'attività amministrativa algoritmica tra etica benthamiana e "black box": il futuro ha un cuore antico?*, in *Contr. impr.*, 2025, p. 26.

[27] Principio risalente a Montesquie, il quale, tuttavia, ad oggi, si ritiene in gran parte superato, o comunque ridimensionato. È opportuno evidenziare, però, come la dottrina abbia rimarcato che «l'apertura a siffatto utilizzo dei sistemi di intelligenza artificiale non scalfisce la soggezione del giudice soltanto alla legge che caratterizza il nostro ordinamento, dovendo pertanto intendersi finora solo come ausilio all'attività del giudice nella ricerca di precedenti rilevanti per la decisione, al fine di garantirne il maggior grado di conoscenza possibile» (E. TAGLIASACCHI, *op. loc. ult. cit.*).

[28] Idea avanzata da G.A. AUTUORI, *Giudizio tributario e intelligenza artificiale: un matrimonio (im)perfetto?*, in *Dir. prat. trib. int.*, 2025, p. 734.

[29] G. GORLA, *Diritto comparato e diritto comune europeo*, Milano, 1981, p. 263. Tra l'altro, la considerazione che la norma, nella sua formulazione testuale, raramente fornisca la risposta precisa alla singola questione, è di fatto recepita dallo stesso art. 12 delle Preleggi, che contempla, accanto al criterio dell'interpretazione letterale, anche quello dell'interpretazione teleologica, oltre che l'*analogia legis* e, da ultimo, l'*analogia iuris* (G.A. AUTUORI, *op. cit*., p. 736).

un giudice e di una legge vivente che non si oppongono, ma sono ben conciliati con la presenza fisiologica di lacune nella protezione legale. O ancora, un magistrato che assume un ruolo propulsivo e propositivo nel comparto giustizia [30] in veste di un vero e proprio «regista del giusto processo» [31].

L'identificazione del contenuto effettivo della norma da applicare al caso concreto passa inevitabilmente attraverso un'attività interpretativa alla quale, in primo luogo, partecipano i giudici di merito, non limitandosi a selezionare e applicare le norme, ma fornendo una lettura in conformità con la Costituzione [32], il diritto dell'Unione europea e il diritto internazionale. Ma non è tutto. Oltre a individuare la norma da applicare tra una moltitudine di fonti che non si adattano perfettamente in modo coerente (il teorico del diritto Herbert L.A. Hart ha descritto la legge come un «tessuto aperto» [33]) e a interpretarla, la giurisprudenza deve anche, ai fini decisionali, selezionare e valutare una moltitudine di elementi e fatti rilevanti *a priori* indefinibili.

Questi ultimi, che possono essere qualificati come profili di specificità del caso individuale contestato o come elementi essenziali dello stesso, influenzano e calibrano necessariamente la decisione giudiziaria in misura tale che una macchina in grado di riconoscerli *a priori* non è concretamente concepibile. Al contrario, i medesimi sono necessariamente propri dell'attività tipicamente umana del distinguere, che a sua volta è bilanciata dall'obbligo di un ragionamento che sia ragionevole ed esaustivo.

Pertanto, se la decisione non è l'applicazione meccanica di regole, un processo decisionale basato sulla generalizzazione statistica delle decisioni precedenti garantirebbe ciò che potrebbe essere chiamato la «decisione del giudice medio», ma certamente non quella del «giudice giusto». In questo senso, è emblematica la convinzione espressa da Alan Turing, secondo cui una macchina non sarà mai in grado, tra le altre cose, di «avere iniziativa, avere senso dell'umorismo, distinguere il bene dal male, commettere errori, innamorarsi, gustare le fragole con la panna» [34]. O si pensi alle parole di Piero Calamandrei

[30] C. ROMEO, *Perplesse valutazioni su giustizia predittiva: vicende di mutazione epocale e riflessi nell'ambito del lavoro*, in *Mass. giur. lav.*, 2023, p. 529.

[31] Termine utilizzato da C. ROMEO, *Processo del lavoro e i nuovi stilemi tra legge e giurisprudenza*, *op. cit.*, p. 8.

[32] Il problema è ben enfatizzato da G. RICOTTI-M.E. CASTIGLIA, *Intelligenza artificiale e fisco: come cambia il rapporto tra Amministrazione finanziaria e contribuente*, in *Dir. prat. trib.*, 2025, p. 2222.

[33] H.L.A. HART, *Il concetto di diritto*, Torino, 2002 (*The Concept of Law*, Oxford, 1961).

[34] A. TURING, *Macchine calcolatrici e intelligenza* (1950), in G. LOLLI, *Intelligenza meccanica*, Torino, 1994, p. 139.

che scrisse «(n)oi non sappiamo più che farci dei giudici di Montesquieu, 'êtres inanimés' fatti di pura logica. Vogliamo giudici con l'anima, giudici engagés, che sappiano portare con vigile impegno umano il grande peso di questa immane responsabilità che è il rendere giustizia» [35].

Si può, dunque, riassumere il tutto proponendo un'intuizione di un noto giurista, secondo il quale «ad oggi, il modello adottato per creare motori di ricerca è quello statistico-giurisprudenziale. Tale modello, però, anche se di agevole implementazione, sconta dei limiti applicativi, considerato che, nel nostro sistema, i precedenti sono privi di efficacia vincolante nei confronti del giudice, che, ex art. 101, comma 2, Cost., è soggetto solo alla legge. Il giudice, infatti, non è una macchina calcolatrice, ma un uomo vivo e la funzione di applicare la legge al caso di specie rappresenta 'un'operazione di sintesi, che si compie a caldo, misteriosamente, nel crogiuolo sigillato dello spirito'» [36].

Da tutto quanto detto consegue che un decisore meccanico nasconde il pericolo non solo di una banalizzazione delle decisioni, ma anche di perdere la dimensione emotiva e preziosa del giudizio che, poggiando sulle innumerevoli sfumature delle idee e delle emozioni umane, ha «paradossalmente il suo punto di forza nell'imperfezione» [37]. Ad oggi, il «comune sentire» [38] è ancora che nessuna macchina potrà mai sostituire la sensibilità, l'intuizione e la capacità decisionale che, invece, appartengono esclusivamente all'essere umano. L'umanità del giudice, infatti, deve essere intesa come valore aggiunto e non surrogabile, e rimane, quindi, a caposaldo di ogni soluzione del confronto giudiziario, «residuo mistico e spirituale dell'intelligenza del soggetto» [39], nella primazia dell'intelligenza individuale che ancora oggi mantiene l'egemonia su quella artificiale. Il rischio è, pertanto, di avere una giustizia che sia "disumanizzata". Una simile prospettiva evoca scenari che ricordano da vicino l'angoscia kaf-

[35] P. CALAMANDREI, *Opere giuridiche*, I, *Problemi generali del diritto e del processo*, Roma, 2019, p. 650.

[36] C. ROMEO, *Giustizia predittiva e processo del lavoro: machine learning vs. ragionamento giuridico*, in *Lav. giur.*, 2025, p. 564, il quale a sua volta cita espressamente le parole di P. CALAMANDREI, *Lo storico e il giudice*, in *Riv. dir. proc. civ.*, 1939, p. 123.

[37] G. ZACCARIA, *Figure del giudicare: calcolabilità, precedenti, decisione robotica*, in *Riv. dir. civ.*, 2020, p. 2.

[38] L'espressione è tratta dal libro del cardinale C.M. MARTINI, *Il comune sentire*, Milano, 2011, sintetizzato nella seguente frase «per il servizio della verità essere pronto ad amare le avversità».

[39] C. CANULLO, C*hi decide? Intelligenza artificiale e trasformazioni del soggetto nella riflessione filosofica*, in E. CALZOLAIO, *La decisione nel prisma dell'intelligenza artificiale*, Padova, 2020, p. 25 s.

kiana del Processo, dove Josef Kafka si trova smarrito dinanzi ad una macchina giudiziaria impersonale e inaccessibile [40].

Inoltre, un sistema fondato sull'idea del giudice automatico, che decide sulla base dell'analisi dei precedenti giudiziari (oltre, naturalmente, ai dati del caso e alle norme di legge), inevitabilmente porta a una standardizzazione delle decisioni giudiziarie, impedendo così al sistema stesso di progredire ed evolversi in linea con il cambiamento del contesto di riferimento [41].

Attraverso il ricorso a sistemi di giustizia predittiva, se declinata ovviamente sino alle sue estreme conseguenze di mero riporto di una casistica precedente senza dare conto della specificità del singolo caso concreto, viene smentito il ruolo centrale del giudice nel processo [42]. Autorevole dottrina considera questo meccanismo come un paradosso e, in particolare, quello di una radicale rivoluzione nell'approccio ermeneutico del giudice-interprete alla fattispecie concreta, limitandosi quest'ultimo ad un mero compito di «taglia-incolla» [43] che poterebbe essere giustificato solo se trattasi di cause seriali. A ben vedere, tuttavia, pure in tali casi rimane saldo il principio contenuto nella norma di cui all'art. 111 Cost. che stabilisce che «tutti i provvedimenti giurisdizionali devono essere motivati».

Si tenga, d'altronde, ben presente come un abuso riguardo l'utilizzo della giustizia predittiva, in luogo di un accertamento istruttorio e di una valutazione del caso concreto, conduce ad un'operazione interpretativa nella quale risulta azzerato il potere del giudice da un'opacità algoritmica decisamente criptica e unidirezionale [44].

---

[40] F. KAFKA, *Der Prozess*, Berlin, 1925.

[41] Riguardo all'argomento a cui si riferisce, è molto interessante il confronto tra la legge e la macchina di Galton sviluppata da C. BARBARO, *Uso dell'intelligenza artificiale nei sistemi giudiziari: verso la definizione di principi etici condivisi a livello europeo?*, in *Quest. giust.*, 2018.

[42] Anche se vi sono autori che mostrano un atteggiamento di favore nei confronti dell'evoluzione tecnologica. G. CORASANITI, *Il ruolo del giudice e le nuove frontiere informatiche*, in *Il Sole 24 Ore*, 19 settembre 2024, p. 17, ad esempio, ha osservato come «l'enigma, lungi dall'essere stato risolto, si presenta ancora attuale e al centro di un ampio dibattito: la prevedibilità delle decisioni, infatti, è messa in crisi dalla disgregazione del sistema delle fonti, dal veloce susseguirsi delle norme e dal repentino cambiamento degli indirizzi giurisprudenziali, fattori questi che rendono arduo l'agire dell'interprete e l'attività del giudicante". Quest'ultimo A., tuttavia, «ritiene di poter riporre grande fiducia nella tecnologia e, in particolare, nell'intelligenza artificiale, considerata da molti strumento in grado di rispondere a quell'esigenza di certezza del diritto e, più in generale, di uguaglianza».

[43] Termine utilizzato da C. ROMEO, *Processo del lavoro e i nuovi stilemi tra legge e giurisprudenza*, cit., p. 8.

[44] Invero, è possibile consentire ad un giudice l'utilizzo di algoritmi di giustizia predittiva solo se tali strumenti sono comprensibili ovverosia spiegabili nella loro logica, in modo tale che possano entrare a far parte del più ampio impianto motivazionale che sorregge la decisione. Tutto

Senza dimenticare, peraltro, come sia ancor oggi sconosciuto all'ordinamento italiano la logica del precedente, propria dei sistemi di *Common Law*, la quale solamente potrebbe legittimare una decisione che tragga fondamento su statuizioni su fattispecie passate, probabilmente in conformità alla giurisprudenza maggioritaria [45]. Il tutto con evidenti difficoltà nell'ipotesi in cui il giudice si dovesse trovare a decidere su fattispecie nuove con paradigmi datati o quantomeno non adeguati [46].

Accogliendo una tale logica, ahimè, il giudice non sarebbe più libero, bensì diverrebbe «obbligato» dall'insieme delle decisioni precedenti e, pertanto, come efficacemente affermato da parte della dottrina, «il passato delle decisioni giudiziarie già prese, peserebbe come un macigno immodificabile sulle sue decisioni future» [47].

Sarebbe, pertanto, opportuno operare, nonostante l'utilizzo degli strumenti algoritmici, un c.d. *prospective over ruling* [48], il quale, sia pure adottato nel rispetto del principio di precauzione elaborato dalla giurisprudenza delle Sezioni Unite e al riparo di effetti pregiudizievoli di mutamenti imprevedibili, potrebbe rappresentare un rimedio al pericolo di stagnazione delle decisioni «adagiate» sui precedenti, dunque contro una deleteria sterilizzazione del processo. È, infatti, ad oggi, sufficientemente consolidato l'orientamento per cui nel nostro ordinamento giuridico l'interpretazione offerta dalla giurisprudenza possa mutare in ragione di diverse e nuove esigenze. Vi è, d'altronde, la necessità che vi sia un costante adattamento del diritto vivente ai continui mutamenti sopravvenuti negli scenari del mercato e della società [49].

---

ciò in dispregio ad una pronuncia che dovrebbe contenere la valutazione delle prove offerte *ex* art. 116 c.p.c. nel segno di una direzione del processo e delle singole udienze sottoposti al governo del giudice dal momento che infinite possono essere le forme di bilanciamento dei principi e degli interessi contrapposti, a seconda delle peculiarità del caso concreto.

[45] G. RICOTTI-M.E. CASTIGLIA, *op. cit*., p. 2237.

[46] Come acutamente osservato da V. MASTROIACOVO, *Processo tributario e giustizia predittiva*, in F. GIULIANI (a cura di), *Tax & Digital*, Bari, 2024, p. 226, «proprio nel tentativo di definire il diritto, pur in un quadro legislativo – in ipotesi – immutato, riemerge l'insoluto dilemma tra interpretazione e produzione della norma, tra la visione del positivismo sillogico-deduttivo di un ordinamento ricurvo al suo interno e privo di lacune e quella dello storicismo che, appunto contrapponendosi al riduzionismo giuridico, esalata la fattualità del diritto facendo leva sull'emersione inesauribile di nuove fattispecie ad opera esclusiva dell'interprete e del giudice».

[47] A. SIMONCINI, *op. cit*., p. 415.

[48] Il *prospective overruling* è una tecnica giurisprudenziale con cui un giudice cambia l'interpretazione di una norma rispetto al passato, ma stabilisce che la nuova interpretazione valga solo per il futuro, e non per i casi precedenti.

[49] Sul punto si rimanda a uno dei padri fondatori della teoria generale del diritto: cfr. E. BETTI,

La mancata modifica delle decisioni giudiziali, inoltre, comporterebbe un ingiusto pregiudizio al diritto degli individui di ottenere una decisione equa, che costituisce l'espressione della funzione primaria del diritto, ovvero quella di fornire risposte alle esigenze umane regolando le relazioni in modo corrispondente alle esigenze sociali del particolare momento storico. Sarebbe, quindi, un immobilismo legale dovuto all'imposizione di un vincolo – anche meccanico – di conformità automatica. E, si badi, anche i sistemi di *Common Law* tendono a mitigare questo tipo di rischio, seppur storicamente si basano (come in precedenza detto) sulla regola dello *stare decisis*, stimolando un dialogo figurato tra il giudice del caso attuale e quelli del caso passato con l'intento di verificare l'applicabilità dello stesso principio decisionale anche al secondo caso deciso.

Allo stesso tempo, l'accettazione di un sistema puramente predittivo non solo comprometterebbe – in vista della cristallizzazione – la Giurisprudenza nel suo complesso, ma avrebbe anche effetti negativi sui singoli magistrati in termini di imparzialità e deresponsabilizzazione.

Questi, in primo luogo, si affiderebbero in maniera pressoché totalizzante sulle decisioni della Intelligenza Artificiale con tutto ciò che ne consegue in termini di *over-reliance* [50]. In altri termini, i giudici si adagerebbero sulle risposte fornite dall'algoritmo tanto per superficialità quanto per motivi di «giustizia difensiva». In particolare, la soluzione algoritmica svilupperebbe un meccanismo di «adattamento» nella misura in cui il giudice pigro non andrebbe oltre il semplice risultato fornito dalla macchina, accettandolo in modo tendenzialmente passivo e senza procedere all'analisi della questione alla luce delle particolarità del singolo caso. Anche definito da alcuni come l'effetto «performativo» della previsione tecnologica, la quale non ha più un valore indicativo, ma quasi prescrittivo, per il decisore umano. Il rischio sarebbe quello di affidarsi ad esse in maniera acritica, fino a svuotare di significato lo *ius dicere* e la funzione stessa del giudice umano [51]. Ne deriverebbe una trasformazione radicale

L'interpretazione della legge e degli atti giuridici (teoria generale e dogmatica), Milano, 1949, p. 48 ss.

[50] Per *over-reliance* (o *automation bias*) si intende la fiducia eccessiva dell'essere umano nelle decisioni prese da un algoritmo, tale da portarlo ad accettarle automaticamente senza verificarle criticamente. Cfr. sul tema, tra i molti, M. BIANCO, *L'intelligenza artificiale nei sistemi ADR: opportunità e presidi*, Conciliatore Bancario Finanziario, Assemblea – incontro con gli Associati, Roma, 17 luglio 2025.

[51] Alcuni autori hanno utilizzato il verbo «catturare» per esprimere l'effetto che un utilizzo acritico dell'intelligenza artificiale potrebbe comportare. Cfr. A. SIMONCINI, *op. cit*., p. 409. Sul punto (ovverosia quanto l'inserimento di un sistema automatizzato all'interno di una procedura decisionale finisca per catturare la decisione stessa), si vedano le recentissime pronunce Schufa Holding della Corte Europea di Giustizia (*Case C-634/21 and Joined Cases C-26/22 and C-*

dell'equilibrio costituzionale dei poteri: non soltanto la fine dell'indipendenza giudiziaria, così come oggi intendiamo, ma persino l'affievolimento dell'idea di una giurisdizione quale potere autonomo dello Stato di diritto [52].

Allo stesso tempo, d'altra parte, una decisione contraria a una previsione algoritmica implicherebbe l'assunzione di un rischio e, di conseguenza, una maggiore responsabilità e pressione. Non sembra irrealistico immaginare, come correttamente segnalato da alcune dottrine, che «i giudici possano essere riluttanti ad assumersi questo onere, in particolare nei sistemi in cui il loro mandato non è permanente, ma soggetto a votazione popolare, o in quelli in cui la loro responsabilità personale (disciplinare, civile o addirittura penale) può essere applicata, soprattutto se le garanzie previste dalla legge in materia disciplinare sono insufficienti» [53]. E, forse, ancora peggio – a livello ideale – sarebbero le pratiche di valutazione, anche dette di «valutazione sociale», ovvero il monitoraggio e la classificazione di giudici o avvocati in base alle decisioni prese o al comportamento.

Tuttavia, i rischi associati allo sviluppo tecnologico nel campo giudiziario non si limitano a quelli sopra menzionati, ma riguardano molti altri aspetti.

Infatti, i dati inseriti nella macchina possono produrre effetti negativi sulla decisione e, in generale, un deterioramento del sistema giudiziario. A ben vedere, invero, i medesimi non solo contribuiscono a produrre quel fenomeno di cristallizzazione del diritto a cui si è fatto riferimento, ma, cosa ancora più grave, la selezione degli stessi dati e delle loro correlazioni può consacrare possibili preconcetti [54]. In effetti, l'intero percorso dalla raccolta dei dati ai risultati è influenzato dal giudizio individuale, poiché è necessario non solo scegliere il tipo di elementi da archiviare, ma anche stabilire i loro requisiti, il numero mi-

64/22) del 7 dicembre 2023, in cui per la prima volta si applica l'art. 22 del GDPR, che vieta decisioni esclusivamente automatizzate; in queste pronunce – che non vertono su decisioni di natura giudiziaria – si assimila alla decisione interamente automatizzata, la decisione umana presa sulla base di una valutazione automatizzata, laddove, cioè, sia estremamente improbabile che il decisore umano si discosti da tale valutazione; cfr. F. PALMIOTTO, *Op-Ed: «'Scoring' for Data Protection Rights: The Court of Justice's First Judgment on Article 22 GDPR (Case C-634/21 and Joined Cases C-26/22 and C-64/22)»*, in *EULAWLive*, 2024, (https://eulawlive.com).

[52] Sul punto G. GIOIA, *Intelligenza artificiale e indipendenza della magistratura: un equilibrio tra trasparenza e controllo*, in *Riv. dir. proc.*, 2025, p. 1338.

[53] C. BARBARO, *op. loc. cit.*

[54] A. GARAPON-J. LASSEGUE, *Justice digitale. Révolution graphique et rupture anthropologique*, Paris, 2018, p. 257 ss.; L. MALDONATO, *Algoritmi predittivi e discrezionalità del giudice: una nuova sfida per la giustizia penale*, in *Dir. pen. cont.*, 2019, II, p. 407; J. NIEVA-FENOLL, *Intelligenza artificiale e processo*, Torino, 2019, p. 65.

nimo da considerare qualitativamente e quantitativamente sufficiente per essere significativo, e così via. L'intero gruppo di queste ultime attività è svolto, a sua volta, da un soggetto umano che, quindi, suggerisce alla macchina di afferrare ciò che vuole che afferri. Il tutto a tal punto che parte della dottrina ha affermato che l'algoritmo «è ontologicamente condizionato dal sistema di valori e dalle intenzioni di chi ne commissiona la creazione e/o di chi la crea» [55]. A questo proposito, il caso riguardante il famoso caso Loomis, deciso dalla Corte Suprema del Wisconsin in merito alla valutazione del pericolo di recidiva ai fini della quantificazione della pena per una persona già riconosciuta colpevole, è stato molto discusso. In quell'occasione, è stato dimostrato, come frutto di un pregiudizio della macchina stessa, che l'impegno e la piena remissione della decisione al sistema COMPAS rivelava l'esistenza di una discriminazione principalmente razziale [56]. Questo caso è un chiaro esempio di come, pertanto, si potrebbe determinare un «pregiudizio di ritorno» [57] nell'ipotesi in cui i *bias*, immessi anche inconsapevolmente dal programmatore nella fase di avvio del *machine learning*, si riflettano sui destinatari che ricevono gli *output* recependoli senza metterli in discussione, ma confidando nella correttezza della risultanza algoritmica. Una parte della dottrina, sul punto, si è spinta sino ad affermare che trema «al pensiero che l'errore in tali indicazioni possa indirizzare la macchina per una via sbagliata» [58].

L'opacità decisionale che discenderebbe da tutto quanto detto avrebbe come conseguenza la circostanza per la quale uno dei fondamentali passaggi del ragionamento giurisdizionale, quello della individuazione del fatto e della sua prova, sembrerebbe diventare uno dei profili più critici. Si potrebbe pensare, ad esempio, a un sistema predittivo di valutazione delle prove, all'interno del quale si debbano inserire gli elementi a disposizione e che poi possa dire se un fatto è da ritenere provato o meno, applicando, in caso negativo, in automatico, la regola dell'onere della prova. A prima vista, un'ipotesi del genere sembra richiamare i

[55] S. SIGNORATO, *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo*, in *Riv. dir. proc*., 2020, p. 614. Lo stesso concetto è trattato anche da T. ALESCI, *Processo algoritmico e contrasto alla violenza di genere. Punti di contatto,* in *Dir. pen. proc.*, 2025, p. 1330.

[56] Un altro caso interessante ha riguardato la società Amazon, la quale è stata condannata per condotta discriminatoria per aver utilizzato un algoritmo nella selezione del personale che, sulla base dello studio delle assunzioni degli ultimi dieci anni, era portato automaticamente a preferire candidati di sesso maschile.

[57] Espressione utilizzata da T. ALESCI, *op. cit*., p. 1341.

[58] I. PAGNI, *A proposito di un libro di Giovanni Verde («Sul potere giudiziale e sull'inganno dei concetti»)*, in *Riv. dir. proc.*, 2024, p. 929.

medioevali cataloghi sul valore delle prove, ove si aveva una pesatura delle prove, secondo criteri legati alla mentalità del tempo: così era stabilito, ad esempio, che la testimonianza di un uomo valesse come quella di 39 donne, che quella di un uomo di chiesa (un *testis clericus*), valesse ancora di più, o ancora che un teste soltanto non avesse valore (*testis unus testis nullus*) [59].

Senza dimenticare come un sistema di intelligenza artificiale porterebbe con sé la difficoltà di rendere comprensibile ai diretti interessati la *ratio* sottesa alla decisione e il ragionamento giuridico alla base della stessa. Sul punto specifico è recentemente intervenuta la Corte di Giustizia dell'UE, che ha affermato in maniera chiara e inequivocabile un principio fondamentale nell'era dell'intelligenza artificiale e dei processi decisionali automatizzati. Si tratta della sentenza relativa alla causa C-203/22 (Corte UE 27 febbraio 2025, C-203/22) che espressamente ha riconosciuto il c.d. «diritto alla spiegazione» delle decisioni algoritmiche in capo agli interessati, i quali, quindi, «potranno rivolgersi ad aziende ed enti pubblici chiedendo una descrizione chiara e comprensibile delle logiche ad esse sottese, ottenendo informazioni significative sulle variabili utilizzate e il loro peso nel determinare il risultato finale» [60]. Orbene, la decisione riportata rappresenta, senza dubbio alcuno, un punto di svolta per la regolamentazione dell'uso degli algoritmi nei processi decisionali e impone ai titolari del trattamento obblighi stringenti in termini di trasparenza e chiarezza dei meccanismi a presidio delle scelte adottate.

[59] Esempi riportati da C. OLIVA, *Predictive justice and factual judgement*, in *Dir. prat. trib. int.*, 2025, p. 799. L'A. aggiunge che «Forse l'ipotesi non è da escludere a priori. Per l'intelligenza umana è spesso difficile catalogare e tenere a mente, contemporaneamente, diversi profili, sicché il giudice, spesso, si trova a decidere sulla base della prova che più lo ha colpito o che ha esaminato per ultima. In questo senso, la possibilità di fornire all'algoritmo i vari elementi probatori può esser d'ausilio per una visione completa ed accurata del materiale probatorio. Naturalmente, poi, occorrerà l'intelligenza umana, che scelga, che valuti e che, tutto considerato, decida se la prova è raggiunta oppure no. E, ove la prova non sia raggiunta, l'algoritmo potrà nuovamente intervenire e applicare la regola dell'onere della prova».

[60] Così testualmente N. SANDON, *Il diritto alla spiegazione: la CGUE rafforza le garanzie sulla trasparenza algoritmica*, in *Compliance*, 24, 2025, p. 7. «Le implicazioni della pronuncia sono molteplici e di non poco momento. Da un lato, le aziende che fanno ricorso - direttamente o tramite fornitori terzi - a processi decisionali automatizzati (quali banche, assicurazioni, aziende di ricerca e selezione del personale, *delivery*, ecc.), anche e soprattutto su base algoritmica, saranno chiamate ad adeguare le proprie policy di gestione dei dati, prevedendo informative più dettagliate e strumenti e processi idonei a fornire, su richiesta degli interessati, le informazioni richieste dalla Corte in un formato chiaro e comprensibile. Dall'altro, viene posto un – ulteriore – accento sul ruolo delle autorità di controllo: queste saranno chiamate in prima persona a vigilare sull'effettivo bilanciamento tra il diritto alla trasparenza e la tutela di eventuali segreti commerciali, che non potranno più essere invocati in modo assoluto per preservare un velo di opacità sulle dinamiche e le variabili coinvolte nei processi decisionali automatizzati».

Senza dimenticare uno dei principali e più dibattuti problemi connessi al tema dell'utilizzo dell'intelligenza artificiale non solo nel comparto giustizia, ma in tutti gli ambiti del sociale, ovvero la questione inerente alla responsabilità civile. Ci si è più volte chiesti, invero, chi debba rispondere civilmente nel caso in cui la decisione algoritmica si dovesse dimostrare produttiva di danni nei confronti dei propri destinatari. In dottrina sono state avanzate le più eterogenee soluzioni che, senza alcuna pretesa di esaustività e in termini assolutamente riassuntivi, si riportano di seguito. Fra le varie ipotesi, invero, è dato riscontrare:

i. L'opinione di coloro che ritengono che si possa riconoscere una forma di soggettività giuridica in capo direttamente alla macchina, sicché ad essa stessa potrebbe essere direttamente imputata una eventuale lesione subita da soggetti terzi. Accogliendo una tale ricostruzione, si tratta, dunque, di valutare la condotta di sistemi in grado di apprendere autonomamente e di decidere sulla base dei dati raccolti. Un primo parametro utilizzabile è quello relativo alle capacità umane. Le capacità dell'uomo potrebbero costituire – almeno nel primo periodo – una soglia minima di diligenza richiesta al sistema operativo.

ii. coloro i quali, invece, sostengono che non sia in alcun modo necessario procedere alla stipula di nuove normative *ad hoc*, essendo sufficiente una rivisitazione e una interpretazione estensiva della disciplina codicistica. All'interno di questo filone, poi, è possibile distinguere a seconda della norma che si ritiene essere applicabile alle lesioni provocate dall'AI. Oltre, ovviamente, a una piccola parte di dottrina incline ad utilizzare la norma generale di responsabilità aquiliana, la restante parte ha preferito di volta in volta richiamare le diverse disposizioni aventi ad oggetto le più varie forme di responsabilità aggravata legislativamente previste;

iii. l'idea, ad oggi, prevalente, di imputare la responsabilità da danno prodotto dalla macchina in capo (solidamente o *uti singulis*) ai soggetti a questa più prossimi, ovvero il produttore, il venditore, etc.

Oltre a quanto già detto in merito al rapporto tra macchina e «proprietario» in termini di discrezione e soggettività, c'è un altro aspetto che va sottolineato. Ovvero la circostanza per la quale, essendo strumenti tecnologici prodotti all'interno di aziende private, i medesimi sono protetti da diritti di privativa o segreto industriale, con tutto ciò che questo comporta in termini di rischio di opacità nella selezione dei dati e degli algoritmi di calcolo e valutazione.

Infine, un'altra vittima del Giudice Automatico sarebbe l'idea giuridica del «giudizio degli uguali», espressamente sancita dalla *Magna Charta Libertatum* dal 1215 e, a partire da oggi, sancita dall'articolo 111 della Costituzione. Non ci sarebbero più due soggetti posti sullo stesso livello e in competizione rispetto al processo avversariale per ottenere ciò che rivendicavano davanti a un terzo soggetto imparziale. Al contrario, ci sarebbe una macchina che, non dando la

possibilità di combattere a parità di condizioni, deciderebbe in modo indipendente e al di sopra dei contendenti.

5. Il diritto penale

Passiamo ora rapidamente ad analizzare i diversi campi del sistema giustizia [61] e, più nello specifico, il settore penale. Come più di un autore ha sottolineato «fondato è il rischio che l'accertamento penale degradi in un semplice esperimento da laboratorio, dove l'apporto umano diventa sempre più raro. Eppure, il sistema penale è un sistema personocentrico e personologico, pensato per l'uomo ed affidato al giudizio dell'uomo» [62]. Del resto, la procedura penale, rappresentando la disciplina dei modi dell'accertamento e della decisione finale [63], è una materia del diritto che si nutre di principi e regole che umanizzano il giudizio penale in ogni settore e forma.

Gli esempi in questo senso sono innumerevoli. Si pensi, ad esempio, alle circostanze attenuanti generiche, disciplinate dall'art. 62-*bis* del Codice Penale. Queste, in particolare, consistono in una categoria di attenuanti non tipizzate, che il giudice può riconoscere discrezionalmente quando ritiene che, alla luce delle circostanze del caso concreto e della personalità dell'imputato, il fatto meriti un trattamento sanzionatorio meno severo [64].

Proprio per questa loro natura, le circostanze generiche rappresentano un importante meccanismo di personalizzazione della decisione giudiziaria. Il diritto penale, infatti, non può limitarsi ad applicare la pena in modo automatico sulla base della sola fattispecie astratta di reato, ma deve tener conto delle caratteristiche concrete del fatto e della persona che lo ha commesso. Attraverso questo istituto il giudice può valorizzare elementi come il comportamento processuale dell'imputato, il grado di colpevolezza, la condotta successiva al reato, il contesto personale e sociale o qualsiasi altra circostanza che renda il fatto meno grave rispetto al modello tipico.

---

[61] M. COLACURCI, *Quale diritto penale dell'IA? Alcune riflessioni a partire dalla proposta di regolamento dell'Unione Europea*, in *Ius*, 2023, pp. 353-378.

[62] Così V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in U. RUFFOLO, *Intelligenza Artificiale - Il diritto, i diritti, l'etica*, Milano, 2020, p. 547.

[63] In questi termini G. RICCIO, *Ragionando su intelligenza artificiale e processo penale*, in *Arch. pen. web*, 3, 2019, p. 2.

[64] A differenza delle attenuanti specifiche, che sono espressamente previste dalla legge per determinate situazioni, le circostanze generiche hanno una struttura elastica e lasciano al giudice un ampio margine di valutazione.

In estrema sintesi, le circostanze generiche costituiscono uno strumento attraverso il quale il giudice può individualizzare la risposta punitiva, garantendo che la decisione giudiziaria sia il risultato non di un'applicazione meccanica della norma, ma di una valutazione complessiva e personalizzata del caso concreto, evitando così statuizioni eccessivamente rigide o standardizzate e favorendo, al contrario, una pena proporzionata alla reale gravità del fatto e alla personalità del reo.

Non meno rilevante appare il giudizio di bilanciamento delle circostanze di cui all'art 69 c.p., che consiste in un giudizio discrezionale complesso, che include valutazioni morali e di proporzionalità difficili da formalizzare e da far assumere a un algoritmo.

Si pensi ancora all'istituto della sospensione condizionale della pena, per la cui applicazione il giudice valuta se il condannato probabilmente non commetterà altri reati. Trattasi, in particolare, di una previsione sul comportamento futuro del reo che si basa su aspetti personalistici - quali l'atteggiamento dell'imputato in aula, il suo percorso personale o il contesto sociale – difficilmente riducibili a dati numerici affidabili.

Concludendo questa breve rassegna, merita un accenno anche il giudizio di prognosi sulla recidiva, all'interno del quale entrano in gioco intuizione giudiziaria, esperienza e valutazione umana del rischio, non solo statistiche. Si tratta di una stima circa la probabilità che una persona condannata torni a commettere reati in futuro e, pertanto, si sostanzia non in una previsione matematica, ma in un giudizio prognostico, cioè orientato al futuro, basato su una serie di elementi concreti della persona e del fatto.

Un algoritmo, a tali fini, potrebbe usare dati statistici, come l'età, i precedenti o il tipo di reato, ma faticherebbe a cogliere fattori fondamentali, tra i quali figurano l'autenticità del pentimento, l'evoluzione della personalità o l'atteggiamento durante il processo. È, infatti, fuori discussione che l'intelligenza artificiale guarda con difficoltà a eventuali valutazioni empatiche e morali, quali, ad esempio, quelle con riguardo alla credibilità delle persone o all'interpretazione dei motivi a delinquere e del pentimento del reo.

Esiste, però, anche una prospettiva meno radicale, che riconosce nell'intelligenza artificiale un utile strumento di pianificazione del lavoro giudiziario [65], in linea con le esigenze di efficienza della macchina giudiziaria. Come si dirà meglio più avanti, a venire in rilievo è una IA che si ponga a supporto della figura del giudice in carne e ossa, fornendo a questo degli utili strumenti per la decisione.

[65] A. SCALFATI, *IA e processo penale: prospettive d'impiego e livelli di rischio*, in *Proc. pen. Giust*., 2024, p. 1405.

Anche il settore penale, pertanto, può trovare un utile giovamento tramite l'utilizzo di una predittività attenuata, la quale non sovrasti, ma affianchi l'essere umano.

6. Il diritto tributario

In molti si sono domandati se un sistema di intelligenza artificiale sarebbe in grado di essere applicato anche nel settore tributario [66]. E seppur è vero che parte della dottrina ha preso spunto dall'assunto per cui il giudizio tributario si presti specialmente, in ragione delle sue caratteristiche tipiche, all'implementazione di questi sistemi [67], la verità è che le criticità fin qui sollevate valgono ovviamente anche per l'applicazione della giustizia predittiva in questo settore [68].

Va primariamente sottolineato, in questo caso, che la riflessione giurisprudenziale deve tenere conto non solo dei principi generali e dello *ius positum*, ma ha anche la necessità di indagare la *ratio* economica della norma.

Occorre allora passare a vagliare alcuni aspetti peculiari del diritto tributario, i quali influiscono anche sulla possibile applicazione dei sistemi di intelligenza artificiale in questo particolare settore.

Una prima peculiarità da considerare è che le varie tipologie di algoritmi utilizzati per la giustizia predittiva, nel caso del diritto tributario, sarebbero addestrati utilizzando base dati che contengono non solo le norme e la giurisprudenza, ma anche la prassi dell'Amministrazione Finanziaria, come atti di indirizzo, circolari o interpelli, ovvero di documenti mediante i quali l'organo amministrativo indica al giudice le risposte che essa si attenderebbe qualora fosse coinvolta in una controversia. È, pertanto, evidente il rischio, in questo caso, che si perda la terzietà del giudice, quella prerogativa che consiste nel tenere sullo stesso piano le ragioni del contribuente e dell'amministrazione, basando la propria decisione su elementi non "di parte".

Sotto un diverso punto di analisi, ulteriore elemento che emerge con forza è, senza ombra di dubbio, la marcata serialità che connota il contenzioso tributario

[66] G. RICOTTI-M.E. CASTIGLIA, *op. cit*., p. 2222 s.

[67] Si fa, ad esempio, riferimento a G.A. AUTUORI, *op. cit*., p. 734. Dello stesso avviso L.T. MCCARTY, *Reflections on "Taxman": An Experiment in Artificial Intelligence and Legal Reasoning*, in *Harvard LR*, 1977, pp. 837-893; R. CORDEIRO GUERRA, *L'intelligenza artificiale nel prisma del diritto tributario*, in *Dir. prat. trib*., 2020, p. 3; B. KUZNIACKI, *The Artificial Intelligence Tax Treaty Assistant: Decoding the Principal Purpose Test*, in *BIT*, 2018, p. 1 s.

[68] Non bisogna, tuttavia, dimenticare i numerosi benefici che il diritto tributario può trarre dall'utilizzo di sistemi di intelligenza artificiale. Sul punto cfr. G.A. AUTUORI, *op. cit*., p. 740 s.

anche definito «contenzioso di massa» [69]. Non si tratta, sia chiaro, di un dato meramente quantitativo [70], bensì di una caratteristica che attiene alla natura stessa delle controversie fiscali. Nel momento in cui l'Amministrazione finanziaria è chiamata a gestire e verificare decine di milioni di posizioni, quasi sempre mediante procedure automatizzate e standardizzate, è fisiologico che anche il momento giurisdizionale del rapporto tributario assuma caratteri di ripetitività.

Un simile aspetto, certamente, porta con sé inevitabili ricadute in termini di utilizzo della strumentazione tecnologica. Sotto un punto di vista tecnico, infatti, la tendenziale ripetitività delle questioni controverse ben si attaglia al funzionamento di sistemi di IA, che si «nutrono» di dati per allenarsi e per fornire risposte con precisione progressivamente maggiore. Volendo osare con le parole, si potrebbe dire che quello che è, per il giudice umano, un «tedioso déjà-vu», per l'algoritmo è un dato fondamentale da cui apprendere [71].

A questo primo argomento, di carattere ordinamentale, se ne affianca un secondo, relativo in particolar modo alla struttura e alla configurazione delle fattispecie tributarie. Se si accoglie l'idea orami consolidata per cui il processo tributario si presenti come un giudizio di «impugnazione-merito», ne discende che il suo oggetto materiale sia rappresentato dall'atto, emesso dall'Amministrazione finanziaria o altro soggetto, con cui questo avanzi nei confronti del contribuente una pretesa impositiva [72].

Trattasi di una peculiarità di non secondaria rilevanza, dal momento che da essa ben possono farsi discendere almeno due considerazioni, entrambe convergenti nel senso della praticabilità di soluzioni di AI nel processo tributario. *In primis*, è innegabile che quest'ultimo si configura come un giudizio tendenzialmente oggettivo, in quanto incentrato su un *quid* obiettivamente apprezzabile

[69] L'avvento di una vera e propria «fiscalità di massa», con il passaggio dal sistema impositivo del secondo dopoguerra, strutturato su poche imposte reali, verso un sistema di imposizione personale e progressiva che coinvolge milioni di contribuenti, ha determinato, quale suo inevitabile corollario, l'emersione di un contenzioso altrettanto «di massa».

[70] Per quanto i numeri delle controversie pendenti presso le Corti di giustizia tributaria e la Corte di Cassazione siano di per sé eloquenti.

[71] Si pensi, ad esempio, alle controversie sulla spettanza di una detrazione Irpef, moltiplicate nell'ordine delle migliaia, le quali non rappresentano innumerevoli problemi diversi, bensì variazioni infinitesimali e sfumate dello stesso problema, che costituiscono il terreno perfetto per affinare sistemi che migliorano in accuratezza proprio quando hanno a disposizione *datasets* amplissimi su cui esercitarsi.

[72] Ad essere dedotto in giudizio non è, pertanto, né un rapporto giuridico, come avviene nel processo civile (che, non a caso, nasce per arbitrare relazioni intersoggettive tra privati), né un fatto, come avviene nel diritto penale (che mira ad accertare la rilevanza penale di una data condotta), bensì un atto giuridico, di carattere autoritativo e di natura documentale.

nella sua esistenza e nel suo modo di essere. In questione, infatti, non è né l'esistenza, né il contenuto dell'atto impositivo, bensì la legittimità e la fondatezza della pretesa con esso recata. Peraltro, salvo ipotesi circoscritte, in cui assumano rilievo concetti elastici e sfumati [73], la fattispecie tributaria è normalmente composta da elementi obiettivamente apprezzabili, che, come tali, non richiedono uno sforzo valutativo o comunque discrezionale quale quello imposto parimenti al giudice civile ed al giudice penale [74].

In secondo luogo, e di conseguenza a quanto detto sinora, tale oggettività si riflette inevitabilmente anche sul versante probatorio. Stante la tendenziale indisponibilità del rapporto impositivo, è esclusa nel processo tributario la possibilità di ricorrere sia al giuramento, che alla prova testimoniale. Ne consegue che l'unico strumento a disposizione delle parti è rappresentato dalla prova documentale, la cui formazione è peraltro imposta, o comunque prevista, dalla stessa normativa fiscale. Trassi, pertanto, di una forma di limitazione probatoria, da tempo oggetto di aspre critiche da larga parte della dottrina [75], la quale, tuttavia configura paradossalmente un vantaggio nell'ottica dell'integrazione dell'AI nel giudizio tributario. A venire in rilievo, infatti, sono, nella maggior parte dei casi, prove documentali precostituite e che presentano una struttura ed un contenuto standardizzato [76]. Per tale ragione, è evidente che esse ben si prestano ad essere «date in pasto» a sistemi di *machine learning*.

---

[73] Basti pensare alla nozione di inerenza, o di abuso del diritto.

[74] A. GAMBINO, *I sette vizi capitali dei giudici-robot (tra blockchain e AI)*, in *Dir. merc. tecn.*, 2018 (https://www.dimt.it), secondo cui il «quinto rischio» della giustizia robotica risiede proprio nella difficoltà per l'AI di riempire di significato quelle clausole che, «come finestre aperte sulla vita dei consociati, sono strettamente collegate alla lettura tutta contingente e transitoria dell'interprete magistrato».

[75] *Ex multis*, senza alcuna pretesa di esaustività, A. GIOVANNINI, *Riflessioni in margine all'oggetto della domanda nel processo tributario*, in *Riv. dir. trib.*, 1998, p. 1; R. CORDEIRO GUERRA, *La tutela – processuale e procedurale – del contribuente sottoposto a sanzioni nella giurisprudenza della Corte Europea dei Diritti Umani*, in F. BILANCIA-C. CALIFANO-L. DEL FEDERICO-P. PUOTI (a cura di), *Convenzione Europea dei Diritti dell'Uomo e giustizia tributaria italiana*, Torino, 2014, p. 230 ss.; M. GREGGI, *Giusto processo e diritto tributario europeo: la prova testimoniale nell'applicazione della CEDU (il caso Jussila)*, in *Rass. trib.*, 2007, p. 254 ss.; F.P. LUISO, *Il principio del contraddittorio e l'istruttoria nel processo amministrativo e tributario*, in *Dir. proc. amm.*, 2000, p. 343; P. RUSSO, *Il divieto della prova testimoniale nel processo tributario: un residuato storico che resiste all'usura del tempo*, in *Rass. trib.*, 2000, p. 568 ss.; F. TESAURO, *Istituzioni di diritto tributario: profili di compatibilità con il diritto europeo*, in *Riv. dir. trib.*, 2019, p. 125 ss.

[76] Cfr. F. FARRI, *La giustizia predittiva in materia tributaria*, in *Riv. tel. dir. trib.*, 12 ottobre 2022, p. 5. L'Autore, pur dubitando dell'impiego dell'intelligenza artificiale in sistemi di c.d. giustizia predittiva, ne ammette comunque l'utilità per tutti quei casi in cui il ragionamento giu-

Preso atto degli enormi benefici che il settore tributario potrebbe trarre dall'utilizzo dei sistemi di IA nella risoluzione delle controversie, non bisogna sottovalutare i rischi connaturati a un simile mutamento tecnologico. In altre parole, questa maggiore predisposizione del diritto tributario non può tramutarsi in una sorta di «lasciapassare» per l'introduzione di sistemi decisionali intelligenti, imponendo anzi una cautela diversa e più stringente rispetto a quella che sarebbe normalmente richiesta per altri settori. A rendere particolarmente delicata l'introduzione dell'IA nel giudizio tributario appare essere, innanzitutto, l'asimmetria tra i soggetti coinvolti, la quale, a ben vedere, non si esaurisce nel momento procedimentale, ma si riflette anche proprio sulla fase giurisdizionale, dove si sostanzia nella disparità di risorse probatorie e difensive a disposizione delle parti [77]. È, infatti, nota la «imparità delle parti» che contraddistingue il rapporto d'imposta, ove alla posizione del contribuente, animata da interessi individuali e privati, si contrappone il Fisco, dotato di poteri autoritativi e portatore di interessi di carattere pubblicistico. Proprio per tali motivi, il giudice tributario è chiamato ad un compito delicatissimo, che si risolve in una ricerca costante di un equilibrio tra l'interesse dello Stato alla corretta esazione del tributo e l'esigenza fondamentale di tutelare i diritti del contribuente. Tale aspetto non esclude sicuramente che possa farsi ricorso a strumenti intelligenti nella risoluzione delle controversie, ma, sicuramente, richiede che ogni soluzione algoritmica sia strutturata in modo tale da non esacerbare tale disparità, dovendo piuttosto tendere al pieno rispetto delle garanzie sancite dall'ordinamento [78].

---

ridico si presenti organizzato secondo schemi intellegibili dalla macchina. Nella generalità di questi casi, la premessa maggiore del sillogismo giuridico è estremamente semplificata e standardizzabile, la premessa minore altrettanto e la sussunzione si limita generalmente a un riscontro estrinseco, che non si traduce in un ragionamento giuridico complesso. Tutte considerazioni che, a nostro avviso, possono essere ragionevolmente estese anche ad una buona parte dei giudizi che popolano i ruoli delle Corti di Giustizia Tributaria.

[77] Basti pensare, ad esempio, alle banche dati e ai *databases* ad esclusiva disposizione dell'Amministrazione finanziaria, unita al potere unilaterale di questa di richiedere informazioni ulteriori al contribuente o di sollecitare ad altre Amministrazioni, anche straniere, la condivisione di dati rilevanti. Di questo avviso, *ex multis*, R. CORDEIRO GUERRA, *L'onere della prova nel transfer price*, in F. ANELLI-A. BRIGUGLIO-A. CHIZZINI-M. DE POLI-E. GRAGNOLI-M. ORLANDI-L. TOSI (a cura di), *L'onere della prova*, Padova, 2024, p. 1412; S. MULEO, *Il principio europeo dell'effettività della tutela e gli anacronismi delle presunzioni legali tributarie alla luce dei potenziamenti dei poteri istruttori dell'amministrazione finanziaria*, in *Riv. trim. dir. trib.*, 2012, p. 685 ss.

[78] Si veda sul tema L. DEL FEDERICO, *Sui diritti dei contribuenti nell'era della trasformazione digitale*, in *Dir. prat. trib. int.*, 2024, p. 71; G. PITRUZZELLA, *Dati fiscali e diritti fondamentali*, *ivi*, 2022, p. 1.

7. Il diritto di famiglia

Anche il diritto di famiglia è confluito nella lente della giustizia predittiva, con specifico riguardo ai criteri per la determinazione degli effetti economici della crisi familiare. È infatti, noto come essi lascino ampio spazio alla discrezionalità del giudice, (teoricamente) in vista della tutela dei soggetti deboli. Principi e clausole generali elastiche, mutevoli anche alla luce dei *revirement* giurisprudenziali, come quello in materia di assegno divorzile, poteri officiosi così ampi riconosciuti al giudice, nell'interesse dei minori e del coniuge debole, hanno prodotto situazioni disomogenee nella concreta determinazione degli effetti economici della crisi coniugale, percepite spesso come inique. È stato evidenziato come, seppur le disparità siano giustificate da un'esigenza di giustizia del caso concreto, «la costante flessione del numero dei matrimoni, da un lato, e la crescente incidenza dei divorzi, dall'altro, testimoniano che il problema dell'estensione e della durata degli obblighi economici conseguenti al divorzio ha conseguito una diffusione sociale assai significativa», e «la diffusa percezione di incertezza che caratterizza ogni singolo giudizio si riflette sul piano collettivo» [79].

Si è posto il problema di ricorrere a strumenti di intelligenza artificiale al fine di determinare le conseguenze economiche della crisi del matrimonio e conferire un più elevato livello di precisione, prevedibilità e obiettività alle disposizioni legislative e alle loro interpretazioni giurisprudenziali consolidatesi negli orientamenti elaborati dalla giurisprudenza di legittimità.

Ove si accogliesse l'idea di una «giustizia famigliare predittiva» si dovrebbe anche accettare che essa vada ad operare in base ad un approccio interdisciplinare tra diritto, statistica ed economia.

Si tratterebbe di ammettere anche in questo settore «"diritto razionale formale" di Max Weber, totalmente svincolato dalla nozione di giustizia, "un diritto calcolabile – ossia fondato su fattispecie normative, giudizio di sussunzione e metodologia ermeneutica – ... in grado di offrire solo la (relativa) precisione del proprio fondamento: non giustizia né altro che non sia la positività delle scelte, la coerenza tra norma e decisione, la possibilità di contare su ciò che verrà". Un diritto senza valori, ridotto a mera forma: una procedura prevedibile, che offre sicurezza ai traffici economici» [80].

Esiste, nel nostro ordinamento, un importante precedente, in materia di danno alla persona. All'inizio degli anni Duemila una pronuncia del Tribunale

[79] Così E. AL MUREDEN, *Solidarietà post-coniugale e intelligenza artificiale tra clausole generali e sistemi di calcolo scientifici*, in *AJI*, 2022, p. 31.

[80] Così A. MORACE PINELLI, *L'intelligenza artificiale nella crisi della famiglia. Alla ricerca di un diritto calcolabile giusto*, in *Riv. dir. civ.*, 2025, p. 286.

di Firenze, nel determinare l'*an* e il *quantum* di un assegno di mantenimento in favore di un coniuge separato e dei figli, ha fatto ricorso ad un metodo di calcolo econometrico dimostrandone l'utilizzabilità nel settore del diritto di famiglia [81]. Da allora sono stati sviluppati numerosi metodi di calcolo obiettivi [82] in grado di circostanziare le clausole generali che governano gli effetti economici della crisi familiare, secondo un approccio improntato al perseguimento di una effettiva sostenibilità economica supportata da evidenze scientifiche. Si andrebbe così a garantire un elevato grado di prevedibilità delle decisioni e, al contempo, «un incentivo alla soluzione concordata degli assetti economici che scaturiscono dalla crisi della coppia» [83], ovvero alla giustizia preventiva e consensuale, con indubbi vantaggi in termini di deflazione del contenzioso.

Il ruolo del giudice non può, tuttavia, essere esautorato dall'impiego degli algoritmi e ciò trova conferma nel recente regolamento europeo sull'intelligenza artificiale. È indubbio, infatti, che l'AI Act fornisca ulteriori strumenti per dare vita ad una giustizia predittiva, in campo familiare, rispettosa dei diritti fondamentali dei soggetti coinvolti. Un diritto che, nella determinazione degli effetti patrimoniali della crisi familiare, possa avvalersi dell'ausilio della tecnologia, ma pur sempre «controllata» dal giudice, chiamato comunque ad «orientare verso giustizia la formazione del diritto "effettivo"», ricercando il significato della norma che attui il più equo bilanciamento degli interessi in conflitto, nella consapevolezza – posta in luce da un illustre Maestro – che il diritto costituisce «uno strumento per la costruzione di una società giusta» [84].

8. Al di là del bene e del male

Alla luce di tutto quanto detto sinora, è possibile constatare che, ad oggi, si contendono il campo due diversi modelli di giustizia predittiva: la giustizia automatizzata, nella quale l'AI assume un ruolo decisionale autonomo, e la giu-

[81] Trib. Firenze, 3 ottobre 2007, in *Fam. dir.*, 2008, p. 39, con nota di E. AL MUREDEN, *Tenore di vita e assegni di mantenimento tra diritto ed econometria*.

[82] Per la loro disamina, cfr. E. AL MUREDEN, *Solidarietà post-coniugale e intelligenza artificiale*, cit., p. 22, spec. nota 5; B. CAMBRIA, *Crisi della famiglia e quantificazione degli assegni: i software esistenti e le prospettive offerte dall'intelligenza artificiale*, in *Jus civ.*, 2024, p. 14.

[83] E. AL MUREDEN, *op. ult. cit.*, p. 21; G. BERTOLI, *"Il metodo Palermo"*, in AA.VV., *Gli assegni di mantenimento tra disciplina legale e intelligenza artificiale*, Torino, 2020, p. 243 ss.

[84] C.M. BIANCA, *Il diritto privato e la sfida sociale*, in *Iustitia*, 1980, p. 213 ss., il quale osserva che «non basta un diritto giusto per avere una società giusta perché la giustizia è un bene che va conquistato giorno per giorno nella misura in cui si riesce a dare risposta concreta alle esigenze di libertà e solidarietà della persona».

stizia assistita. Per ciò che concerne il primo caso, in Cina, ad esempio, i tribunali di Hangzhou, Pechino e Guangzhou hanno sperimentato piattaforme ODR in cui gli algoritmi decidono controversie di bassa complessità (contratti e-commerce, liti di consumo), con controllo giudiziale minimo [85]. È d'uopo osservare, tuttavia, che la possibilità di introdurre un giudice-macchina «non significa affatto che questo esito sia auspicabile, né che debba essere accettato come un esito inevitabile dello sviluppo tecnologico» [86].

Dal canto suo, il modello assistito, prevalente in Europa e negli ordinamenti occidentali, attribuisce all'IA una funzione ancillare. Si pensi all'esperienza francese dove strumenti come *Predictice* [87] sono stati pensati per assistere gli avvocati nella valutazione delle probabilità di accoglimento di una domanda, del *quantum* medio dei risarcimenti e della durata delle controversie [88]. Esperimenti conclusi come *CaseCrunch* in UK [89] o *ROSS Intelligence*

[85] Tali esperienze, apprezzate per la evidente efficienza, sollevano però critiche per la compressione dei diritti processuali, l'opacità delle motivazioni e il rischio di ridurre il giudice a mero validatore dell'output algoritmico (Cfr., ad es., W. JI, *The Change of Judicial Power in China in the Era of Artificial Intelligence*, in *Asian JLS*, 2020, p. 515 ss.). Un passo intermedio è il progetto estone del giudice *robot* per cause sotto i 7.000 €, che pur basato su un solido contesto digitale solleva analoghi interrogativi di indipendenza (E. NIILER, *Can AI Be a Fair Judge in Court? Estonia Thinks So*, in *Wired*, 25 marzo 2019; V. TANGERMANN, *Estonia is Building A "Robot Judge" to Help Clear Legal Backlog*, in *Futurism*, 25 marzo 2019).

[86] C. CONTESSA, *Dare giustizia e riconoscere diritti non può coincidere con l'utilizzo di algoritmi*, in *Il Sole 24 Ore*, 6 febbraio 2024, p. 42.

[87] La sperimentazione è stata avviata dalle Corti d'Appello di Douai e Rennes, ma non ha funzionato. Cfr. T. COUSTET, *L'utilisation de l'outil Predictice déēoit la cour d'appel de Rennes*, in *Dalloz act.*, 16 ottobre 2017; L. GERARD-D. MOUGENOT, *Justice robotisée et droits fondamentaux*, in *Le juge et l'algorithme: juges augmentés ou justice diminuée?*, Bruxelles, 2019, p. 13 ss.

[88] Il progetto francese *DataJust*, avviato dal Ministero della Giustizia nel 2020 per sviluppare un algoritmo di quantificazione standardizzata del danno civile, è stato abbandonato nel 2022 per timori legati a opacità, limitazione della discrezionalità giudiziaria e rischio di *barémisation* delle decisioni. Sul quale, v. G. DE PASQUALE, *La giustizia predittiva in Francia: il trattamento DataJust*, in *Judicium*, 26 novembre 2020 e L. VIAUT, *DataJust ou l'analyse algorithmique de la jurisprudence*, in *Rev. just. Act.*, 2021, p. 52 ss.

[89] Nel 2017 la startup britannica ha organizzato una competizione nota come *Lawyer Challenge*, nella quale oltre cento avvocati hanno dovuto prevedere l'esito di controversie relative al PPI *mis-selling*, in parallelo con il sistema predittivo di intelligenza artificiale *CaseCruncher Alpha*. L'IA ha raggiunto un'accuratezza dell'86,6%, a fronte del 62,3% ottenuto dai legali. Si trattava, tuttavia, di un esperimento condotto in condizioni controllate, con quesiti binari e su casi già decisi, che non ha trovato applicazione diretta nei tribunali. Il caso è rimasto comunque un riferimento simbolico nel dibattito sul potenziale e sui limiti dell'uso dell'AI in ambito giudiziario.

in US [90] hanno mostrato risultati impressionanti, ma non privi di criticità in termini di affidabilità ed errori sistematici. Merita, infine, menzione anche l'esperienza brasiliana, dove, in nome dell'efficienza, sono stati introdotti sistemi di IA incaricati di svolgere una funzione di filtro nell'accesso ai tribunali superiori [91].

Nonostante le molteplici applicazioni pratiche che si stanno sviluppando in tutte le parti del mondo, però, ad oggi, si è ancora indotti inevitabilmente ad ammettere, non senza una nota di sottile sollievo, che le prestazioni dell'intelligenza artificiale applicata al processo sono comparativamente più deboli e imperfette rispetto ad altri campi. Nonostante ciò, sarebbe un grave errore non sfruttare il suo grande potenziale per supportare e migliorare il duro lavoro della magistratura. Infatti, la «giustizia predittiva» è solo una delle molte applicazioni dell'intelligenza artificiale, ma ve ne sono altre, molto interessanti, che vengono sperimentate in alcuni paesi europei.

Volendone menzionarne alcuni, si pensi all'uso che viene fatto in Austria per la lettura rapida, la classificazione e l'attribuzione di atti, ricorsi e documenti alle sezioni di cancelleria competenti; o all'uso di *chatbot* (letteralmente: conversazione con un robot) per indirizzare il cittadino verso una soluzione alternativa della controversia, come in Lettonia.

Prendendo nota di quanto è stato detto in termini di potenzialità e rischi di un nuovo utilizzo degli strumenti di intelligenza artificiale nel campo del diritto, si ritiene di concludere con una considerazione che, anche prima di essere conclusiva, potrebbe e dovrebbe essere intesa come un aspetto preliminare e

[90] *ROSS Intelligence* era una piattaforma di ricerca giuridica basata su IA, che consentiva di interrogare precedenti e opinioni in linguaggio naturale. Ha cessato le attività dopo la causa Thomson Reuters v. Ross Intelligence, nella quale il tribunale ha ritenuto che l'uso non autorizzato di *headnotes* (una specie di massime redatte da esperti) di Westlaw per addestrare l'algoritmo non rientrasse nel fair use del *copyright*. Il caso è diventato un riferimento centrale nel dibattito su *copyright* e dati di *training* per l'AI in ambito legale. A proposito della raccolta dei dati, va citato anche un altro caso, quello del Tribunale di Amburgo, Kneschke/LAION (sentenza 27 settembre 2024, 310 O 227/23): prima decisione europea significativa che valuta se opere protette, reperibili online, possano essere riprodotte per l'addestramento di *dataset* di IA sotto eccezioni quali quella di *text* and *data mining* della Direttiva DSM 2019/790/ UE. Il Tribunale ha ritenuto che, nel caso concreto, l'eccezione non fosse applicabile alla raccolta e al download preventivo delle immagini effettuati con *scraping web*, soprattutto in presenza di una dichiarazione di riserva (*opt-out*) da parte del titolare dei diritti. Tuttavia, ha riconosciuto che la creazione di *dataset* per fini di addestramento può qualificarsi come attività di ricerca scientifica ai sensi dell'art. 3 della stessa Direttiva.

[91] Cfr. F. HARTMANN PEIXOTO, *Projeto Victor: Relato do Desenvolvimento da Inteligência Artificial na Repercussão Geral do Supremo Tribunal Federal*, in *Rev. br. intel. artif. dir.*, 2020, p. 219 s.; A.J. DE OLIVEIRA TEIXEIRA-D. FERREIRA, *Entre acesso à justiça, publicidade e contraditório: a questão da explicabilidade nos processos decisórios dos sistemas Victor e Vitória no STF*, in *Seqüência*, 2024, p. 1 s.

fondamentale per un'evoluzione ordinaria in questa direzione. Le sfide poste dallo sviluppo tecnologico possono essere affrontate in modo più efficace solo attraverso una fusione di conoscenze. A questo proposito, come sottolineato da giuristi e studiosi più autorevoli, la creazione di *team* multidisciplinari composti da magistrati, ricercatori in scienze sociali e informatica sarebbe vantaggiosa sia nella fase di elaborazione che in quella di pilotaggio e implementazione delle soluzioni proposte dalla scienza. A questo scopo, tuttavia, è necessario, a monte, che, attraverso l'insegnamento, gli studenti vengano indirizzati a un uso accorto della tecnologia, a prediligere la conoscenza sulla mera informazione, a saper sistemare la tumultuosa creatività della giurisprudenza e la profluvie degli interventi del legislatore senza farsene schiacciare.

Tutto quanto sostenuto nei paragrafi che precedono sembrerebbe porre la giustizia predittiva in una posizione di conflitto con le fondamenta del nostro sistema giudiziario. Ci sarebbe un tacito abbandono del ruolo del giudice, in netto contrasto con le caratteristiche di naturalità e pre-costituzionalità sancite dall'articolo 25, paragrafo 1 della Costituzione. Così come sembrerebbe esserci l'annullamento dell'obbligo, stabilito dall'articolo 111, paragrafo 6 della Costituzione, di un controllo della motivazione da parte delle persone in nome delle quali «la giustizia è amministrata», la negazione della soggezione giurisdizionale «solo alla legge» *ex* articolo 101 della Costituzione e la frustrazione del processo avversariale [92]. A ben vedere, pertanto, il ruolo del giudice non può essere esautorato dall'impiego degli algoritmi, giacché la decisione non può prescindere, nel nostro ordinamento, dalla motivazione [93]. Più in generale, la sostituzione di un giudice con un robot è preclusa dall'art. 25 Cost., che sancisce il principio secondo cui nessuno può essere distolto dal proprio giudice naturale, elidendo una simile novità, nella sostanza, l'esercizio del diritto di difesa riconosciuto dall'art. 24 Cost. [94]. È possibile, dunque, immaginare una IA d'ausilio al giudice ma non che lo sostituisca [95], incontrando la decisione automatizzata,

[92] G. FANELLI, *L'impiego dell'intelligenza artificiale nei processi decisori del giudice, tra la disciplina europea e quella del processo civile*, in R. GIORDANO-A. PANZAROLA-M. PROTO-A. POLICE-S. PREZIOSI (a cura di), *Il diritto nell'era digitale. Persona, Mercato, Amministrazione, Giustizia*, Milano, 2022, p. 993; M. LUCIANI, *La decisione giudiziaria robotica*, in A. CARLEO (a cura di), *op. cit.*, p. 86.

[93] R. BICHI, *Intelligenza artificiale e diritto. Intelligenza artificiale tra "calcolabilità" del diritto e tutela dei diritti*, in *Giur. it.*, 2019, p. 1657 ss.

[94] F. AULETTA, *Accertamento del fatto e intelligenza artificiale nel processo civile*, in *Dir. proc. civ. it. comp.*, 2024, p. 922 ss.; A. CARRATTA, *Decisione robotica e valori del processo*, in *Riv. dir. proc. civ.*, 2020, p. 419 ss.

[95] Fortemente contrario alla prospettiva della «decisione robotica giudiziaria», A. CARRATTA,

in questo specifico settore, un necessario limite nell'intervento umano [96]. Bisognerebbe, pertanto, sforzarsi di individuare delle indicazioni adeguate attraverso le quali, nel processo decisionale del giudice, l'interazione uomo-macchina può svolgersi affinché rimanga coerente con i principi dello Stato di diritto e con i valori dell'Unione. «Né solo un giudice macchina né una macchina giudice» [97]: il dialogo alla base della nuova idea di giustizia dovrebbe riflettere l'integrazione cooperativa tra due diverse intelligenze. La tecnologia dovrebbe, infatti, essere sempre concepita come un semplice strumento, mai come fine ultimo.

Per quanto il timore di trovarsi dinnanzi a una giustizia puramente predittiva sia, almeno in parte, giustificato dalla portata rivoluzionaria di tutti i nuovi strumenti intelligenti, si deve evitare ad ogni costo di cadere nell'equivoco secondo cui tale sostituzione sia inevitabile in ogni caso. Questo assunto, che domina in larga parte del discorso pubblico e persino accademico, non a caso è raramente motivato in modo rigoroso, dandosi anzi per scontato che ciò rappresenti l'esito naturale, quasi obbligato, del progresso tecnologico.

Se infatti si accede all'idea per cui l'intelligenza artificiale sia destinata a soppiantare del tutto il giudice nella funzione giurisdizionale, si esclude a priori un'opzione che, in realtà, si presenta come la più realistica e, per certi versi, la più auspicabile. Con ciò si vuole fare riferimento a un paradigma cooperativo, il c.d. "*human in the loop*", in cui l'IA affianca il giudice umano, andando a contribuire ad un generale miglioramento della qualità delle decisioni.

Sarebbe, quindi, opportuno adottare una visione un po' asettica del ruolo che la giustizia predittiva potrebbe essere chiamata a svolgere nell'ambito giudiziario. E sia i legislatori sovranazionali che quelli nazionali si stanno muovendo in questa direzione, ciascuno lavorando per delineare coordinate normative e para-

Decisione robotica e valori del processo, in *Riv. dir. proc.*, 2020, p. 512, apre cautamente ad un impiego degli algoritmi predittivi come «strumenti integrativi dei fatti di causa, una sorta di «ausilio tecnico» di cui il giudice-persona potrebbe servirsi negli stessi termini e con gli stessi limiti di cui può servirsi della tradizionale consulenza tecnica». Cfr. anche N. LETTIERI, *Contro la previsione. Tre argomenti per una critica del calcolo predittivo e del suo uso in ambito giuridic*o, in *Algoritmi ed esperienza giuridica, Ars int.*, 2021, p. 113 ss.

[96] Il tema della dicotomia tra umano e digitale è da tempo al centro dell'attenzione del legislatore europeo. Nella Dichiarazione europea sui diritti e i principi digitali per il decennio digitale COM (2022) 28, è espresso l'intento di elevare le aspirazioni delle persone e la loro libertà di scelta a criterio fondante il modello europeo per la transizione digitale. In particolare, alla salvaguardia della libertà di autodeterminazione del soggetto è dedicato l'intero Capitolo III del documento ove si evidenzia che *«l'intelligenza artificiale dovrebbe fungere da strumento per le persone, con l'obiettivo ultimo di aumentare il benessere umano».* Cfr. sul tema, *ex multis*, A. PUNZI, *AI Act: gli effetti settoriali - La decisione giudiziaria nell'AI Act*, in *Giur. it.*, 2025, p. 448.

[97] R. MESSINETTI, *op. cit*., p. 283.

normative in grado di guidare e regolamentare il comportamento degli operatori, compresi quelli legali, nel settore dell'intelligenza artificiale.

In effetti, la Commissione europea per l'efficienza della giustizia (Cepej) ha condotto uno studio approfondito su questi temi e ha redatto una «Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari» [98] che stabilisce principi sostanziali e metodologici per la produzione e l'implementazione di strumenti algoritmici, applicabili anche all'analisi e all'elaborazione delle decisioni giudiziarie.

Tra i criteri guida vi sono il rispetto dei diritti umani e la non discriminazione da parte delle applicazioni di intelligenza artificiale, attraverso la piena conformità con i diritti garantiti dalla Convenzione europea sui diritti umani e dalla Convenzione del Consiglio d'Europa sulla protezione dei dati personali, dalla fase di progettazione all'applicazione pratica. Non meno importante è il principio di trasparenza, ovvero la necessità di rendere accessibili e comprensibili le tecniche di trattamento dei dati, nonché di autorizzare *audit* esterni da parte di parti indipendenti al fine di individuare eventuali distorsioni.

Il trattamento delle informazioni effettuato dal *machine learning*, infatti, dovrebbe essere effettuato sulla base di dati conoscibili (di cui i soggetti devono essere informati in un linguaggio chiaro e comprensibile) e che sono espressioni concrete delle diverse realtà su cui il giudice è chiamato a decidere. Allo stesso tempo, è fondamentale garantire la possibilità per quest'ultimo di deviare dai risultati algoritmici in considerazione delle specificità del caso particolare.

Tuttavia, un ruolo primario viene assunto dalla disciplina adottata in merito al rapporto tra algoritmo e giudice [99]. Anche di fronte al noto caso Loomis, gli organismi europei ritengono che per i sistemi qualificati «ad alto rischio» (che includerebbero quelli della giustizia predittiva) debba necessariamente esistere una qualche forma di supervisione umana, e mettono in guardia anche dal rischio di un'eccessiva dipendenza da parte degli esseri umani nello sviluppo dell'intelligenza artificiale (il cosiddetto *bias* dell'automazione). In altri termini, nel Preambolo della Carta etica europea sull'uso dell'Intelligenza artificiale a venire in rilievo sarebbe quella che alcuni autori hanno definito come una sorta di riserva di umanità [100].

---

[98] Il primo strumento europeo su questo argomento è stato approvato dal Consiglio d'Europa durante la sessione plenaria CEPEJ del 3-4 dicembre 2018 (disponibile all'indirizzo https://rm.coe.int), facendo riferimento a uno studio condotto dall'University College di Londra (UCL), che ha esaminato 584 decisioni di apprendimento automatico della Corte europea dei diritti dell'uomo con un tasso di successo del 79%.

[99] *Ex multis*, N. LETTIERI, *op. cit*., p. 113 s.

[100] E. TAGLIASACCHI, *op. cit.*, p. 31.

La Carta Etica, sebbene, come inizialmente affermato, non sia un atto giuridico vincolante per gli Stati, ha avuto un effetto pervasivo a livello globale, essendo stata una chiara fonte di ispirazione da più parti. A titolo di esempio, ricordiamo l'adozione della Carta dei diritti dell'IA degli Stati Uniti o, a livello europeo, la Convenzione quadro sull'intelligenza artificiale e i diritti umani, la democrazia e lo stato di diritto, adottata dal Consiglio d'Europa il 17 maggio 2024, il cui scopo è garantire che, nel contesto internazionale, le attività comprese nell'intero ciclo di vita dei sistemi di IA rispettino questi valori [101]. O il recentissimo Regolamento UE 2024/1689 del 13 giugno 2024 sull'intelligenza artificiale, noto anche come AI Act. Pur tuttavia, è opportuno sottolineare che già in epoca precedente il legislatore europeo aveva sentito l'esigenza di contenere i rischi insiti nell'uso dei sistemi algoritmici. Esempi di questo approccio si possono trovare in Art. 22 Reg. n. 2016/679/UE, art. 11 Di. n. 2016/680/UE e art. 8 D.Lgs. n. 51/2018, in cui il cosiddetto «principio di non esclusività» della decisione algoritmica è sancito nella legge sulla privacy. Escludono, cioè, le decisioni supportate esclusivamente da un trattamento automatizzato, compresa la profilazione, poiché le eccezioni – che richiedono comunque adeguate salvaguardie per i diritti, le libertà e gli interessi legittimi degli individui – devono prevedere anche l'intervento umano nel processo decisionale automatizzato relativo alle persone fisiche in modo da non ledere la dignità. Emblematico sul punto è il principio già affermato nell'art. 22 del GDPR secondo cui non è solo vietata «*una decisione basata unicamente sul trattamento automatizzato*», bensì sarebbe prescritta un'efficace supervisione nel corso dell'attività svolta dal sistema di AI e, dunque, un controllo su tutti i passaggi rilevanti ai fini della decisione finale, che deve sempre essere assunta dall'uomo.

Da parte sua, l'Italia ha avviato diversi progetti volti a sperimentare strumenti di giustizia predittiva, sebbene l'uso dell'intelligenza artificiale nel settore giudiziario sia perlopiù rimasto limitato a iniziative locali. Alcuni progetti si basano su semplici banche dati di merito, come i portali delle Corti d'appello di Brescia e Bari, che organizzano la giurisprudenza per aree tematiche senza impiego di algoritmi. Altri, invece, esplorano tecniche più avanzate: a Genova e Pisa l'analisi dei *big data* e il *machine learning* sono utilizzati per stimare l'esito e la durata delle cause; a Catania è stato sviluppato un prototipo per i procedimenti in materia di immigrazione; a Brescia, Genova, Milano e Torino il «Laboratorio delle Sentenze» lavora su strumenti di riassunto automatico e modelli di redazione assistita; a Venezia, infine, è stato avviato un progetto per

[101] Il documento è disponibile presso l'indirizzo istituzionale del *Council of Europe's* (www.co e.int).

fornire previsioni probabilistiche sugli esiti dei giudizi in materia di licenziamenti [102].

Non sono mancate però anche iniziative a più ampio spettro, come dimostra il fatto che l'Italia è stato uno dei primi paesi europei ad agire in conformità con le linee guida sovranazionali e il 20 maggio 2024 ha presentato per l'approvazione parlamentare il disegno di legge n. 1146, intitolato «Disposizioni e delega al Governo in materia di intelligenza artificiale». È evidente la volontà del legislatore nazionale di promuovere un uso responsabile dell'intelligenza artificiale [103].

A tal fine, sembra essenziale mettere al centro del sistema l'uomo e il rispetto dei diritti fondamentali, nonché garantire a) l'uso equo e corretto degli algoritmi; b) la qualità, la sicurezza e la protezione dei dati utilizzati e dei processi con cui vengono elaborati e trattati dai sistemi di intelligenza artificiale; c) la sostenibilità digitale. Per quanto riguarda specificamente il campo della giustizia, essi prendono in considerazione gli artt. 14 e 15, che, in sintesi, riaffermano il concetto di un «essere umano al comando». Questo modo di dire significa che è legittimo utilizzare e trarre vantaggio dai vantaggi derivanti dall'uso di sistemi algoritmici, ma solo al fine di semplificare e assistere la figura del giudice persona fisica, il cui onere e dovere di decisione permane.

È lo stesso art. 14, infatti, a stabilire che «i sistemi di intelligenza artificiale sono utilizzati esclusivamente per l'organizzazione e la semplificazione del lavoro giudiziario, nonché per la ricerca giurisprudenziale e dottrinale». Il Ministero della Giustizia regola l'uso dei sistemi di intelligenza artificiale da parte degli uffici giudiziari ordinari. Per le altre giurisdizioni, l'uso è regolamentato in conformità con le rispettive leggi. La decisione sull'interpretazione della legge, sulla valutazione dei fatti e delle prove e sull'adozione di qualsiasi misura spetta sempre al magistrato. Sono già presenti sul mercato diverse applicazioni di intelligenza artificiale che potrebbero essere utilizzate per perseguire gli scopi previsti dalla legge. Per citarne alcuni, si consideri l'implementazione delle potenzialità di alcuni *database* legali – in particolare, il *Database* del Merito Pubblico (BDP) e il *Database* del Merito Confidenziale (BDR), accessibili solo ai magistrati – utilizzando sistemi di intelligenza artificiale generativa.

Oppure lo sviluppo di strumenti algoritmici che possono identificare, attraverso l'analisi computazionale dei dati, eventuali lacune organizzative – al fine di promuovere una migliore allocazione delle risorse – o strumenti di intelligenza artificiale per la traduzione di documenti, per l'analisi e la revisione di atti.

---

[102] Tutti questi esempi sono riportati da G. GIOIA, *Intelligenza artificiale e indipendenza della magistratura: un equilibrio tra trasparenza e controllo*, *op. cit.*

[103] Il d.d.l. A.S. n. 1146 può essere letto sul sito istituzionale del Senato www.senato.it.

Dispositivi tecnologici a supporto del lavoro del giudice, riducendo l'area di giudizio, garantiscono una maggiore prevedibilità delle decisioni, attraverso l'uso di algoritmi in grado di analizzare grandi quantità di dati giuridici (compresi testi legislativi, precedenti giudiziari e decisioni di tribunali, tribunali di merito, la Corte di legittimità e quelli sovranazionali) e offrono allo stesso giudice strumenti per rafforzare le facoltà razionali e decisionali. È importante notare che ridurre non significa eliminare il necessario controllo e la conseguente responsabilità del giudice.

A sua volta, l'art. 15 espressamente dispone che: «Nei casi di impiego dei sistemi di intelligenza artificiale nell'attività giudiziaria è sempre riservata al magistrato ogni decisione sull'interpretazione e sull'applicazione della legge, sulla valutazione dei fatti e delle prove e sull'adozione dei provvedimenti». Si tratta, come è stato scritto da altri [104], «di una declinazione del c.d. principio antropocentrico, già posto a fondamento generale del regolamento europeo [105]: in virtù di tale principio è la tecnologia ad essere al servizio dell'essere umano e non viceversa, cosicché le decisioni finali devono essere sempre assunte dall'essere umano (e non dalla tecnologia)».

La norma, ancora in itinere e come sopra riportata, rappresenterà – allorché entrerà in vigore – l'approdo ad una centrale modalità di certezza del diritto, perché finisce per confermare uno dei principi fondamentali su cui si fondano gli ordinamenti giuridici democratici e liberali: «il potere del giudice sulla decisione e sull'interpretazione della legge, in uno alla sua discrezionalità sulla valutazione delle prove e dei fatti materiali».

In futuro, quindi, i programmi di giustizia artificiale (compresi quelli predittivi) – se rispettati e garantiti in determinate condizioni – diventeranno un aiuto indispensabile per il lavoro del giudice e svolgeranno una funzione di supporto allo stesso. Con un successivo e significativo controllo umano, avremmo l'integrazione desiderata tra algoritmo e processo decisionale alla luce dei principi fondamentali del nostro diritto. Come già detto in precedenza, l'intelligenza artificiale dovrebbe divenire quindi uno strumento, non un fine [106].

Oltre a tutto quanto detto, si pensi ancora alla nuova legge italiana sull'IA del 23 settembre 2025, n. 132, all'art. 15, comma 3, la quale ha previsto che,

---

[104] F. CRISIANO, *Giustizia e AI: le disposizioni del Ddl 1146/2024 approvato dal Senato il 20 marzo 2025*, in *Norme&Tributi*, 31 marzo 2025.

[105] Cfr. art. 1 Reg. UE 2024/1689 ed i relativi considerando nn. 1, 6, 8, 27, 176

[106] Rimane, pertanto, fermo l'assunto che «*il diritto, basato su regole in continua evoluzione, richiede che l'IA sia uno strumento di supporto, con l'uomo come garante del suo corretto utilizzo*» (Cfr. M. DI BIASE, *Giustizia e AI: serve un equilibrio tra innovazione e tradizione*, in *Guida dir*., 21 dicembre 2024, p. 48).

fino alla compiuta attuazione dell'AI Act, la sperimentazione e l'impiego dei sistemi di intelligenza artificiale negli uffici giudiziari (esclusivamente) ordinari, vanno autorizzati dal Ministero della giustizia. Quest'ultimo, peraltro, sta sperimentando Copilot, uno strumento che andrà a beneficio sia dell'organizzazione dell'ufficio sia dell'attività del giudice.

In quest'ottica, non può non menzionarsi quella parte della dottrina che, traendo le mosse da una visione di sfiducia verso una metamorfosi della giustizia in termini artificiale, rimarca la sola e unica possibilità per la macchina di affiancare, seppur utilmente, la figura umana del giudicante. In particolare, si richiamano le parole di G. Ricotti e di M.E. Castiglia, secondo i quali «Sebbene siano stati lanciati, in Italia così come in altri Stati, progetti pilota di giustizia predittiva, tuttavia – proprio alla luce delle peculiarità e dei rischi sopra evidenziati – i sistemi di Gen-IA esistenti non sono capaci di interpretare una norma, enuclearne la *ratio*, soppesare correttamente i contenuti dei precedenti giurisprudenziali e compiere, in ultima istanza, ragionamenti giuridici complessi. Anche in una prospettiva di ulteriore sviluppo tecnologico dei sistemi di Gen-IA, sarebbe comunque preclusa normativamente la possibilità di giungere a una totale automatizzazione della gestione e decisione dei procedimenti giudiziali da parte di cd. giudici robot» [107].

In conclusione, non sarebbe erroneo affermare come, ancor oggi, la difficoltà di comprendere in concreto le modalità di tale interazione giudice-macchina, così da poterle disciplinare, è l'effetto di una certa impostazione culturale, in particolare della prospettiva angusta di quell'antropocentrismo chiuso che ancora oggi prevale nel dibattito delle scienze sociali in tema di intelligenza artificiale. Come autorevolmente sostenuto da parte della dottrina. «La sfida è cercare di recuperare il tempo perduto (ad esempio ripetendo slogan come quello secondo cui le intelligenze non umane non sono vere intelligenze) e riuscire a comprendere in che modo può svolgersi il dialogo tra il giudice e i sistemi di AI nell'ambito di un processo decisionale che dovrà sempre concludersi con una decisione umana, ma che ad essa può approdare attraverso un percorso nel quale l'intelligenza umana e l'intelligenza artificiale sono chiamate a cooperare» [108].

[107] G. Ricotti-M.E. Castiglia, *op. cit.*, p. 2238. Gli Autori a loro volta richiamano M. Bianco, *op. loc. cit.*

[108] A. Punzi, *op. cit*., p. 452.

9. Conclusione

Ad oggi, i principali dubbi dell'interprete si innestano nel campo di relazioni giuridiche oramai controllate da piattaforme telematiche e applicazioni digitali sempre più evolute e dotate di inquietanti autonomie [109].

L'IA si colloca, dunque, in una zona intermedia tra ausilio tecnico e possibile interferenza con la funzione giurisdizionale. Da un lato, può aumentare efficienza e coerenza; dall'altro, la sua natura statistico-probabilistica incontra limiti strutturali rispetto a principi come la valutazione delle prove, l'interpretazione del diritto e la personalizzazione della decisione. Questa tensione mostra come l'indipendenza del giudice non sia solo una garanzia istituzionale, ma si concretizzi nella possibilità di mantenere il controllo sul processo argomentativo e sulla decisione finale. È su questo terreno che si misura la compatibilità dell'IA con le fonti normative nazionali e sovranazionali a tutela dell'autonomia della giurisdizione.

La vera sfida consiste nel comprendere come questi strumenti, fondati sull'analisi massiva dei dati, possano essere utilizzati senza compromettere l'indipendenza del giudice e i principi costituzionali che presidiano la funzione giurisdizionale.

Concludendo, si può affermare che da un dialogo costruttivo tra giudice umano e *justice machine* possa nascere una «giustizia aumentata» [110] volta non solo ad intercettare gli *errores in procedendo* ma anche al rafforzamento del ragionamento, delle argomentazioni giuridiche e della motivazione [111].

[109] Espressione utilizzata da C. ROMEO, *Processo del lavoro e i nuovi stilemi tra legge e giurisprudenza*, cit., p. 5 s.

[110] Termine proprio di R. MESSINETTI, *op. loc. ult. cit.*

[111] Sul tema è possibile richiamare alcuni scritti di A. PUNZI, *Judge in the Machine. E se fossero le macchine a restituirci l'umanità del giudicare?*, in A. CARLEO (a cura di), *op. cit.*, pp. 305-316; ID., *Decidere in dialogo con le macchine: la sfida della giurisprudenza contemporanea*, in U. SALANITRO (a cura di), *op. cit.*, pp. 261-273.

AI LAW

INTERNATIONAL REVIEW
OF ARTIFICIAL INTELLIGENCE LAW

La responsabilità civile del medico utilizzatore di sistemi di intelligenza artificiale nell'ordinamento giuridico italiano

Mirko Faccioli

Professore associato di Diritto privato
nell'Università degli Studi di Verona

Abstract

(EN): Civil liability for damages caused by healthcare activities carried out using artificial intelligence involves a chain of potential liable subjects, ranging from the physician to the hospital, and extending to those who, in various ways, contribute to the development and distribution of smart medical devices. In this context, the role of the healthcare professional takes on particular significance, as their liability is specifically regulated in the Italian legal system by the so-called Gelli-Bianco Law (Law No. 24/2017) as well as by the provisions of the Civil Code regarding contractual and non-contractual liability. When physicians use artificial intelligence systems, this regulatory framework ultimately intersects with the AI Act and raises complex hermeneutic issues, particularly regarding the reconstruction of the criteria for attributing liability, which this paper aims to examine systematically.

(ES): La responsabilidad civil por los daños causados por la actividad sanitaria realizada mediante el uso de la inteligencia artificial implica una cadena de posibles responsables que van desde el médico hasta la estructura sanitaria, llegando hasta los sujetos que, a título diverso, contribuyen a la fabricación y puesta en circulación de dispositivos médicos inteligentes. En este contexto, cobra especial relevancia la figura del profesional sanitario (professionista sanitario), cuya responsabilidad está regulada en el ordenamiento jurídico italiano específicamente por la denominada Ley Gelli-Bianco (Ley n. 24/2017), además que por las disposiciones del Código civil en materia de responsabilidad contractual y extracontractual. Cuando el médico utiliza sistemas de inteligencia artificial, este conjunto de normas se entrelaza con el AI Act y plantea delicadas cuestiones hermenéuticas, relacionadas sobre todo con la reconstrucción del criterio de imputación de la responsabilidad, de las que el presente artículo se propone ofrecer un examen sistemático.

(FR): La responsabilité civile pour les dommages causés par des activités médicales menées à l'aide de l'intelligence artificielle implique une chaîne de responsables potentiels allant du médecin à l'établissement de santé, jusqu'aux sujets qui, à divers titres, contribuent à la conception et à la mise sur le marché de dispositifs médicaux intelligents. Dans ce contexte, la figure du profession-

nel de santé revêt une importance particulière; la responsabilité de ce dernier est, dans l'ordre juridique italien, spécifiquement réglementée par la loi dite «Gelli-Bianco» (loi n. 24/2017) ainsi que par les dispositions du Code civil en matière de responsabilité contractuelle et extracontractuelle. Lorsque le médecin utilise des systèmes d'intelligence artificielle, cet ensemble de règles s'entremêle avec l'AI Act et soulève des questions herméneutiques délicates, relatives surtout à la reconstruction du critère d'imputation de la responsabilité, dont la présente contribution entend proposer un examen systématique.

(DE): Die zivilrechtliche Haftung für Schäden, die durch medizinische Behandlungen unter Einsatz künstlicher Intelligenz verursacht werden, betrifft eine ganze Kette potenziell haftbarer Akteure, die vom Arzt über die Gesundheitseinrichtung bis hin zu denjenigen reicht, die in unterschiedlicher Funktion an der Entwicklung und dem Vertrieb intelligenter medizinischer Geräte beteiligt sind. In diesem Zusammenhang kommt der Figur des Arztes besondere Bedeutung zu, dessen Haftung im italienischen Rechtssystem speziell durch das sogenannte Gelli-Bianco-Gesetz (Gesetz Nr. 24/2017) sowie durch die Bestimmungen des Zivilgesetzbuches zur vertraglichen und außervertraglichen Haftung geregelt ist. Wenn der Arzt Systeme der künstlichen Intelligenz einsetzt, verflechten sich diese rechtlichen Rahmenbedingungen mit dem AI Act und werfen schwierige hermeneutische Fragen auf, die vor allem die Rekonstruktion der Haftungsgrundlagen betreffen und die in diesem Beitrag systematisch untersucht werden sollen.

(PT): A responsabilidade civil por danos causados por atividades de saúde realizadas com uso da inteligência artificial envolve uma cadeia de potenciais responsaveis. Os responsaveis são o médico, a instituição de saúde e todos os sujeitos que fabricam os dispositivos médicos intelligentes. Neste contexto assume particular relevância a figura do profissional de saúde. A responsabilidade no ordenamento jurídico italiano é regulada pela Lei Gelli-Bianco (Lei n.º 24/2017) e pelas normas do Código Civil sobre a responsabilidade contratual e extracontratual. Quando o médico utiliza sistemas de inteligência artificial, esse conjunto normativo está conectado com regulamento europeu sobre a Inteligência artificial e levanta delicadas questões hermenêuticas sobre a reconstrução do critério de imputação da responsabilidade. O artigo oferece uma análise sistemática sobre o tema.

(IT): La responsabilità civile per danni cagionati da attività sanitaria svolta tramite l'impiego dell'intelligenza artificiale coinvolge una filiera di potenziali responsabili che vanno dal medico alla struttura sanitaria fino ad arrivare ai soggetti che, a vario titolo, concorrono a realizzare e mettere in circolazione dispositivi medici intelligenti. In questo contesto, assume particolare rilievo la figura del professionista sanitario, la cui responsabilità è nell'ordinamento giuridico italiano appositamente disciplinata dalla c.d. Legge Gelli-Bianco (l. n. 24/2017) oltre che dalle previsioni del codice civile in tema di responsabilità contrattuale ed extracontrattuale. Quando il medico utilizza sistemi di intelligenza artificiale, questo plesso normativo finisce per intrecciarsi con l'AI Act e sollevare delicate questioni ermeneutiche, relative soprattutto alla ricostruzione del criterio di imputazione della responsabilità, delle quali il presente contributo intende offrire un esame sistematico.

Sommario: 1. Introduzione: la centralità della figura del medico nella catena della responsabilità civile per danni derivanti da prestazioni mediche eseguite con l'IA. – 2. La responsabilità conseguente all'inadempimento degli obblighi generalmente imposti ai *deployer* dell'IA. – 3. La responsabilità derivante dallo scorretto utilizzo degli

output prodotti dall'IA in ambito sanitario. – 4. L'incidenza dell'utilizzo dell'IA sull'applicazione dell'art. 2236 c.c. in ambito sanitario. – 5. L'inconfigurabilità di una responsabilità oggettiva del medico per i danni cagionati al paziente impiegando l'IA. – 6. L'incidenza dell'utilizzo dell'IA sulla ricostruzione del nesso causale intercorrente tra la condotta del medico e il danno subito dal paziente.

1. Introduzione: la centralità della figura del medico nella catena della responsabilità civile per danni derivanti da prestazioni mediche eseguite con l'IA

Nell'ormai abbondante letteratura stratificatasi sul tema della responsabilità civile ricollegata all'utilizzo dell'intelligenza artificiale (d'ora in avanti, IA) in medicina [1], viene ricorrentemente evidenziato come occorra prestare particolare attenzione a ripartire in maniera appropriata la responsabilità in parola tra i medici, gli enti nosocomiali e quanti a vario titolo concorrono a realizzare e mettere in circolazione dispositivi sanitari intelligenti [2].

In questo quadro, la definizione dei contorni della responsabilità del singolo professionista riveste un'importanza fondamentale alla luce di due ordini di ragioni: innanzitutto perché egli rappresenta, insieme alla struttura sanitaria, il soggetto contro il quale è verosimilmente portato ad agire il paziente danneggiato dalle cure erogate con l'IA [3]; in secondo luogo perché, a prescindere da come si

[1] Per limitarsi alle sole trattazioni monografiche dell'argomento in lingua italiana, v. F. LAGIOIA, *L'intelligenza artificiale in sanità: un'analisi giuridica*, Torino, 2020; G. PASCERI, *Intelligenza artificiale, algoritmo e* machine learning*. La responsabilità del medico e dell'amministrazione sanitaria*, Milano, 2021; A. DI MARTINO, *Intelligenza artificiale e responsabilità civile in ambito sanitario*, Milano, 2022; F. FERRETTI, *Tutela della salute, intelligenza artificiale e gestione del rischio*, Napoli, 2025; E. COLLETTI, *Intelligenza artificiale e attività sanitaria: i nuovi danni*, Bari, 2025.

[2] Cfr., tra gli altri, A. D'ADDA, *Danni «da robot» (specie in ambito sanitario) e pluralità di responsabili tra sistema della responsabilità civile ed iniziative di diritto europeo*, in *Riv. dir. civ.*, 2022, p. 807 ss.; A. FIORENTINI, Machine learning *e dispositivi medici: riflessioni in materia di responsabilità civile*, in *Corr. giur.*, 2021, p. 1260; M. FACCIOLI, *Intelligenza artificiale e responsabilità sanitaria*, in *Nuova giur. civ. comm.*, 2023, p. 735; A. COLARUOTOLO, *Intelligenza artificiale in sanità tra etica, responsabilità civile e consenso informato*, in *Jus civ.*, 2025, p. 282; D. GIANTI, *Quando a curare è un algoritmo: ripensare la responsabilità medica in ottica comparata nell'era dell'intelligenza artificiale*, in R. FAVALE-M.A. MANTOVANI (a cura di), *Le intelligenze artificiali nel sistema della responsabilità civile. Prospettive applicative nel settore sanitario*, Napoli, 2026, p. 195 ss.

[3] Mentre appare assai meno probabile, per ragioni facilmente intuibili, che il paziente decida di indirizzare un'azione risarcitoria nei confronti dei soggetti che hanno a vario titolo partecipato

ritenga preferibile distribuire il carico risarcitorio conseguente lungo la catena dei (potenziali) responsabili sopra menzionati, a dare origine alla vicenda di *malpractice* è sempre l'operatore sanitario che concretamente adopera un sistema di IA nell'esercizio della sua attività.

Anche all'ambito delle cure mediche, difatti, può essere senz'altro riferita l'autorevole opinione di chi, riflettendo sulla responsabilità per danni da IA in una prospettiva di carattere generale, ha osservato che «il diritto civile ha imparato, nel corso del Novecento, che la responsabilità va allocata laddove produce il maggior effetto di riduzione del danno. Ed è qui che entra in gioco la figura dell'utilizzatore: […] il vero "locus" della prevenzione. Chi decide se attivare un algoritmo, in quale contesto operativo, con quali dati, per quali finalità, è l'unico attore che possiede una conoscenza concreta della situazione. È questa conoscenza situata che permette di valutare rischi, adottare cautele, e, se necessario, interrompere l'uso. È quindi naturale che l'utilizzatore sia il perno della responsabilità, non perché sia colpevole, ma perché è l'unico che può realisticamente modificare la condotta in funzione preventiva» [4].

D'altro canto, occorre osservare che il medico detiene la direzione e il controllo della prestazione eseguita con l'ausilio dell'IA, posto che nessuna delle macchine intelligenti impiegate nei più svariati settori della medicina, per quanto avanzata e sofisticata possa essere, è in grado di sostituirsi completamente all'operatore umano e portare a termine in completa autonomia un'intera procedura diagnostica o terapeutica [5].

E del resto, il concreto utilizzo di un sistema di IA che dovesse superare questo limite di carattere tecnico si scontrerebbe, sul piano normativo, con il dettato dell'art. 7, comma 5, l. 23 settembre 2025, n. 132, il quale stabilisce che «i sistemi di intelligenza artificiale in ambito sanitario costituiscono un supporto nei processi di prevenzione, diagnosi, cura e scelta terapeutica, lasciando impregiudicata la decisione, che è sempre rimessa agli esercenti la professione

alla fabbricazione, alla programmazione e alla messa in circolazione dei dispositivi intelligenti impiegati nelle cure: sul punto v., per tutti, F. CEREA, *La responsabilità civile da sistemi "intelligenti"*, Torino, 2025, p. 173 ss.

[4] P.G. MONATERI, *Sulla non responsabilità delle intelligenze artificiali per il danno a terzi*, in *Danno resp.*, 2026, p. 6.

[5] Per maggiori dettagli al riguardo, ormai facilmente reperibili anche all'interno delle trattazioni giuridiche della materia, v., *ex multis*, G. PASCERI, *op. cit.*, p. 45 ss.; A. DI MARTINO, *op. cit.*, p. 9 ss.; G. NATALE, *L'intelligenza artificiale in sanità. Il dialogo necessario tra medicina, etica e diritto*, Milano, 2025, p. 101 ss.; U. COMITE, *L'intelligenza artificiale in sanità: vantaggi e rischi*, Torino, 2025, p. 81 ss.; O. POLLICINO-F.A. SACCHI-N. CONDITI, *Is AI the Perfect Doctor? Artificial Intelligence's Global Impact on the Legal and Policy Boundaries in Healthcare*, Milano, 2025, p. 22 ss.

medica». Attraverso questa previsione, il legislatore italiano ha invero fatto proprio con riguardo a tutte le soluzioni di IA utilizzate in medicina il principio di «sorveglianza» o supervisione umana consacrato – in particolare, ma non esclusivamente – dall'art. 14 *AI Act* (Reg. UE 2024/1689) [6] con specifico riguardo ai sistemi di IA ad alto rischio [7], categoria nella quale appaiono, comunque, destinati a rientrare la gran parte dei sistemi intelligenti impiegati nell'ambito della sanità [8], ovverosia: ai sensi dell'art. 6, par. 1, del regolamento e del suo all. I, i

[6] Oltre che nell'art. 14, ulteriori manifestazioni e corollari del principio di supervisione umana sono infatti presenti in diversi altri luoghi normativi dell'*AI Act*: v., sul punto, N. ZORZI GALGANO, *La sorveglianza umana sull'"alto rischio" nel Regolamento UE 2024/1689 del 13 giugno 2024 nei sistemi di intelligenza artificiale*, in *Contr. impr. Eur.*, 2025, pp. 239 ss., spec. 261 ss.

[7] A. MICHINELLI, *Legge 23 settembre 2025, n. 132. Articolo 7. Uso dell'intelligenza artificiale in ambito sanitario e di disabilità*, in P. PERRI-G. VACIAGO-G. ZICCARDI (a cura di), *Commentario AI Act e legge nazionale sull'intelligenza artificiale. Regolamento (UE) 2024/1689 e legge n. 132/2025*, Milano, 2026, p. 1088 s.; P. GAMBATESA, *L'uso dell'intelligenza artificiale in ambito sanitario. Commento agli artt. 7, 8 e 10*, in M. D'AMICO-F. BALAGUER CALLEJÓN-A. AGUILAR CALAHORRO-P. GAMBATESA-S. DI GIOVANNI (a cura di), *Commentario tematico del Regolamento (UE) 2024/1689 e della legge italiana 23 settembre 2025, n. 132*, Torino, 2026, p. 555 s., il quale evidenzia che «la scelta del legislatore nazionale di esplicitare tale principio in ambito sanitario appare particolarmente significativa, poiché consente di chiarire, in un settore ad alta sensibilità etica e giuridica, che l'IA non è chiamata a sostituire il giudizio clinico, ma a integrarlo. Al tempo stesso, la norma sembra voler prevenire il rischio che l'automazione decisionale alteri l'assetto delle responsabilità professionali, mantenendo ferma la centralità del medico come titolare della decisione e come interlocutore primario del paziente».

[8] Di converso, in ambito sanitario dovrebbe assai difficilmente assumere rilievo la categoria delle pratiche di IA vietate ai sensi dell'art. 5 *AI Act*, il quale non menziona espressamente l'ambito in discorso ed, anzi, attribuisce a «motivi medici» il ruolo di eccezione al divieto, stabilito dalla lett. *f*) della norma, di immettere sul mercato, mettere in servizio o utilizzare sistemi di IA volti a inferire le emozioni di una persona fisica sul luogo di lavoro e negli istituti di istruzione (di «sistemi destinati all'uso terapeutico» parla espressamente, al riguardo, il 44° *considerando* del regolamento in questione). A bandire taluni impieghi dell'IA in ambito sanitario potrebbe al limite essere, in futuro, il disposto della lett. b) dell'articolo in discorso, che menziona i sistemi di IA che sfruttano le vulnerabilità di una persona fisica o di uno specifico gruppo di persone (quali sono i pazienti), dovute all'età, alla disabilità o a una specifica situazione sociale o economica, con l'obiettivo o l'effetto di distorcerne materialmente il comportamento in un modo che provochi o possa ragionevolmente provocare loro un danno significativo (quale senz'altro è il danno alla salute): S. LATTANZI, *Prime riflessioni sull'applicazione del nuovo regolamento sull'intelligenza artificiale al settore sanitario*, in *Quaderni AISDUE*, fasc. spec. n. 2/2024, p. 187 s. A tale riguardo occorre, peraltro, tenere conto pure del 29° *considerando* dell'*AI Act*, nel quale si legge che «i divieti di pratiche manipolative e di sfruttamento di cui al presente regolamento non dovrebbero pregiudicare le pratiche lecite nel contesto di trattamenti medici, quali il trattamento psicologico di una malattia mentale o la riabilitazione fisica, quando tali pratiche sono svolte conformemente al diritto applicabile e alle norme in ambito medico».

sistemi di IA che sono (o sono componenti di sicurezza di) dispositivi medici ai sensi del Reg. UE 2017/745 oppure dispositivi medico-diagnostici in vitro ai sensi del Reg. UE 2017/746, qualora sottoposti a una procedura di valutazione di conformità da parte di un organismo terzo [9]; ai sensi dell'art. 6, par. 2, del regolamento e del suo all. III, i sistemi di IA utilizzati da autorità pubbliche o per conto di autorità pubbliche per valutare l'ammissibilità delle persone fisiche ai servizi di assistenza sanitaria e i sistemi di IA utilizzati per valutare e classificare le chiamate di emergenza o per inviare servizi di emergenza di primo soccorso o per stabilire priorità in merito all'invio di tali servizi, compresa l'assistenza medica, nonché per i sistemi di selezione dei pazienti per quanto concerne l'assistenza sanitaria di emergenza [10].

Né sembra inappropriato considerare, quantomeno in linea di principio, «previsioni aventi […] carattere generale (non dunque necessariamente limitate ai casi di rischio elevato)», quando riferite all'ambito sanitario, le disposizioni sugli obblighi dei fornitori e dei *deployer* dei sistemi di IA ad alto rischio dettate dagli artt. 16-27 dell'*AI Act* [11], tra le quali viene ai nostri fini in rilievo, in particolare, la norma sugli obblighi degli utilizzatori dei sistemi di IA contenuta nell'art. 26 del regolamento, ai sensi della quale i medici e gli enti nosocomiali sono, in estrema sintesi, tenuti a: adottare misure tecniche e organizzative atte a garantire un utilizzo dei sistemi di IA conforme alle istruzioni che li accompagnano; affidare la sorveglianza umana a persone che dispongono della compe-

[9] Sulla complessa opera di coordinamento tra questi due plessi normativi e l'*AI Act*, v. F. GENNARI, *O Complementarity, Where Art Thou? Wading through the Medical Device Regulation and the AI Act Compliance: The case of Software as a Medical Device. A Primer*, in *Bio LJ*, 2024, p. 411 ss.; M. EBERS, *AI Robotics in Healthcare Between the EU Medical Device Regulation and the Artificial Intelligence Act. Gaps and Inconsistencies in the Protection of Patients and Care Recipients*, in *Oslo LR*, 11, 1-2024, p. 1 ss.; S. PALMIERI, *AI Act and MDR: Investigating Regulatory Overlaps, Influences, and Joint Application Issues*, in V.L. RAPOSO (a cura di), *The European Artificial Intelligence Act: Promises and Perils?*, Cham, 2025, p. 339 ss.

[10] Su questi aspetti, cfr. S. LATTANZI, *op. cit.*, p. 188 ss.; F. FERRETTI, *op. cit*., p. 28 ss.; O. POLLICINO-F.A. SACCHI-N. CONDITI, *op. cit.*, p. 116 ss.; C. PERLINGIERI, *Intelligenza artificiale in ambito medico-sanitario: profili di ricostruzione normativa a seguito dell'*AI Act, in *Rass. dir. civ.*, 2024, p. 921 s.; S. DE LUCIA, *I.A. e responsabilità civile in medicina: verso un'innovazione tecnologica sostenibile e umanocentrica*, in *AJI*, 2025, p. 1191 ss.; P. GAMBATESA, *I sistemi di IA ad alto rischio in ambito sanitario. Commento all'Allegato III, par. 5*, in M. D'AMICO-F. BALAGUER CALLEJÓN-A. AGUILAR CALAHORRO-P. GAMBATESA-S. DI GIOVANNI (a cura di), *op. cit*., p. 246 ss.; G. PROIETTI, *Definire l'indefinibile? I sistemi di intelligenza artificiale alla ricerca di un inquadramento sistematico*, in *Contr. impr.*, 2024, p. 917 s.

[11] Così G. DI ROSA, *Robot e responsabilità sanitaria*, in A. ALBANESE-L. GUFFANTI PESENTI-S. TOMMASI (a cura di), *Il diritto privato nell'economia digitale. Tutele giurisdizionali e autorità indipendenti*, Torino, 2025, p. 230.

tenza, della formazione e dell'autorità necessarie, garantendo loro adeguato sostegno, anche in collegamento con il più generale obbligo di alfabetizzazione in materia di IA di cui all'art. 4 *AI Act*; garantire che i dati di *input* siano pertinenti e sufficientemente rappresentativi alla luce della finalità del sistema; monitorare il funzionamento del sistema, sospenderne l'uso e informare i fornitori e le autorità pertinenti in caso di incidenti gravi o di rischi per la salute, la sicurezza o i diritti fondamentali innescati dall'uso conforme alle istruzioni ricevute; conservare i *log* generati automaticamente dal sistema che si trovano sotto il loro controllo per un periodo, adeguato alle finalità del sistema stesso, di almeno sei mesi; informare le persone fisiche interessate qualora utilizzino sistemi di IA che adottano decisioni o assistono nell'adozione di decisioni che le riguardano [12].

Com'è stato efficacemente evidenziato, «la normativa in questione, applicata al settore sanitario, si pone così in linea di continuità con il principio espresso dall'art. 1 della legge Gelli-Bianco, che stabilisce un dovere generale di prevenire e gestire i rischi per la salute dei pazienti attraverso "l'utilizzo appropriato delle risorse strutturali, tecnologiche e organizzative" di cui dispongono le strutture preposte, con il contributo di tutto il personale medico e paramedico» [13]. La prestazione medica eseguita con l'ausilio dell'IA, difatti, rientra a pieno titolo nell'ambito di applicazione della disciplina *in subiecta materia* prevista dalla l. 8 marzo 2017, n. 24, che per il professionista sanitario si traduce, principalmente, nell'assoggettamento all'obbligo di attenersi alle linee guida e alle buone pratiche clinico-assistenziali di cui all'art. 5 della legge e alle disposizioni sulla responsabilità civile per i danni arrecati nell'esercizio della propria attività previste dall'art. 7 della stessa [14].

[12] Riferiscono queste previsioni all'ambito sanitario S. LATTANZI, *op. cit.*, p. 193 ss.; O. POLLICINO-F.A. SACCHI-N. CONDITI, *op. cit.*, p. 121 ss.; H. VON KOLFSCHOOTEN-J. VON OIRSCHOT, *The EU Artificial Intelligence Act (2024): Implications for healthcare*, in *Health policy*, 149, 2024, p. 1 ss.

[13] A. ALBANESE, *Digitalizzazione delle cure e intelligenza artificiale: gli obblighi integrativi strumentali e di protezione nell'adempimento della prestazione sanitaria*, in *Resp. med.*, 2025, p. 308 s. Sul rilievo assunto dal principio di sicurezza delle cure nell'ambito delle prestazioni mediche eseguite con l'IA si sofferma anche M.P. MANTOVANI, *Le intelligenze artificiali nel settore sanitario*, in R. FAVALE-M.A. MANTOVANI (a cura di), *op. cit.*, p. 168 ss.

[14] V., *ex multis*, A. MICHINELLI, *op. cit.*, p. 1089; M. FACCIOLI, *op. cit.*, p. 735; C. PERLINGIERI, *op. cit.*, p. 908 ss.; F.A. SCHURR, *Alcune riflessioni sulla "malpractice tecnologica" e sulla responsabilità sanitaria per l'uso di sistemi di intelligenza artificiale*, in *Resp. med.*, 2022, p. 225 ss.; R. SCOTTI, *La responsabilità civile dei danni cagionati da sistemi di intelligenza artificiale in ambito sanitario*, in *Giust. civ.*, 2024, p. 167 ss.

Sulla scorta di queste ultime, che tra l'altro regolamentano, insieme ai principi codicistici sulla responsabilità contrattuale ed extracontrattuale, un aspetto trascurato dall'*AI Act* [15] e inizialmente affidato ad una proposta di direttiva poi ritirata dalle istituzioni europee [16], il medico può, come noto, alternativamente rispondere a titolo di responsabilità contrattuale o extracontrattuale a seconda che abbia o meno «agito nell'adempimento di obbligazione contrattuale assunta con il paziente». Nell'ambito della nostra indagine, che sarà incentrata sulla va-

[15] La scelta di non disciplinare la responsabilità civile nell'*AI Act* è stata da taluno criticata osservando che «sviluppare un modello di tutela *ex ante*, incentrato sull'analisi del rischio, senza poi elaborare un quadro adeguato e coerente per le ipotesi residuali in cui la cattiva o carente gestione dei rischi sia causa di danni finisce per menomare l'impatto complessivo dell'intervento normativo […], che risulta così un'opera incompiuta in una visione ampia della regolamentazione dell'AI» (A. MANTELERO, *L'AI Act: la risposta del legislatore europeo alle sfide dell'intelligenza artificiale*, in *Acc.*, 2024, p. 197). In senso opposto, altra dottrina ha invece sostenuto che «questo approccio sottende probabilmente la presa d'atto di una funzionalità adeguata dei regimi esistenti, in particolare di quelli che non siano basati sul puro criterio della colpa, a rispondere alle criticità poste dagli illeciti algoritmici. La creazione di un corpo normativo orizzontale, destinato ad essere applicato ogni qualvolta un prodotto basato sull'IA fosse coinvolto nella vicenda che ha causato l'evento, avrebbe viceversa sacrificato i regimi settoriali, i quali hanno invece una maggiore aderenza alle circostanze concrete in cui avvengono certe tipologie di illecito, che ricorrono sempre quale che sia il livello di "intelligenza" e sofisticazione degli strumenti adoperati. In molti casi, inoltre, le attività in cui si origina il danno si svolgeranno in maniera mista, ossia in parte secondo modalità tradizionali, in parte avvalendosi di dispositivi intelligenti: cosicché sottrarre una porzione dell'attività al regime settoriale per applicare le regole orizzontali potrebbe non risolversi in una facilitazione per le vittime, che dovrebbero impegnarsi non soltanto nella dimostrazione degli elementi richiesti ai fini della prova dell'illecito, ma anche, e prima ancora, nella ricerca delle regole di volta in volta applicabili» (E. PALMERINI, *Il problema dei (nuovi) danni negli illeciti algoritmici*, in *Resp. civ. prev.*, 2025, p. 423 s.).

[16] Si tratta, per la precisione, della Proposta di Direttiva del Parlamento europeo e del Consiglio sulla responsabilità extracontrattuale da Intelligenza Artificiale COM(2022) 496 final, che peraltro contemplava soltanto misure volte ad alleggerire l'onere probatorio da assolvere nell'esercizio di un'azione di responsabilità extracontrattuale per colpa allo scopo di ottenere il risarcimento del danno causato da un sistema di IA, senza intervenire sugli aspetti generali della responsabilità civile come determinati dai singoli ordinamenti nazionali, quali la definizione stessa di colpa, il concetto di causalità, le tipologie di pregiudizi risarcibili, la quantificazione del danno, la prescrizione, e così via. In argomento v., *amplius*, A. BERTOLINI, *Intelligenza artificiale e responsabilità civile. Problema, sistema, funzioni*, Bologna, 2024, p. 262 ss.; M. FACCIOLI, *La responsabilità civile per danni cagionati da sistemi di intelligenza artificiale nel prisma dell'onere della prova*, in *Resp. civ. prev.*, 2024, p. 958 ss.; M. GAMBINI, *Nuovi paradigmi della responsabilità civile per l'Intelligenza artificiale*, in *Rass. dir. civ*., 2023, p. 1314 ss.; S. CORSO, *Responsabilità per causalità presunta. Intelligenza artificiale, responsabilità civile e diritto eurounitario*, in *Resp. civ. prev.*, 2025, p. 640 ss.

lutazione della condotta del professionista sanitario che utilizza l'IA, tale variabile non assume peraltro rilievo: secondo una consolidata impostazione, la condotta esigibile dal medico, quale soggetto esercente una professione protetta, va, infatti, sempre valutata alla stregua delle norme dettate dagli artt. 1176, comma 2 e 2236 c.c. nonché dagli articoli della Legge Gelli-Bianco sopra menzionati, tali disposizioni designando un modello di comportamento che, essendo ricollegato all'appartenenza a un determinato *status* professionale, non è destinato a mutare in ragione del fatto che il professionista svolga o meno il proprio operato in esecuzione di un contratto stipulato con il paziente [17]. Occorre considerare, inoltre, che anche contro il medico legato al paziente da un contratto d'opera professionale possono trovare applicazione le norme di cui agli artt. 2043 ss. c.c., pacifico essendo che, nelle ipotesi in cui coesistono le fattispecie dell'inadempimento e dell'illecito civile, come tipicamente avviene nell'ambito sanitario, il danneggiato ha la facoltà di agire, a sua scelta, in via contrattuale o in via aquiliana [18].

Alla luce di questi rilievi e nell'attesa di un (eventuale) intervento del legislatore italiano sul tema della responsabilità civile per danni ricollegati all'IA [19], nelle pagine che seguono la responsabilità del medico sarà pertanto indagata in termini unitari e senza distinguere sul punto, se non quando l'inquadramento della natura della responsabilità in esame possa influenzare la disciplina applicabile e conseguentemente la risoluzione delle questioni di volta in volta esaminate [20].

[17] F. FERRETTI, *op. cit*., p. 58 s.; R. DE MATTEIS, *Colpa medica e inadempimento delle strutture sanitarie*, in *Contr. impr.*, 2015, p. 561; G. MONTANARI VERGALLO, *La nuova responsabilità medica dopo la riforma Gelli-Bianco*, Roma, 2017, p. 105.

[18] V., per tutti, C.M. BIANCA, *Diritto civile*, 5, *La responsabilità*, III ed., Milano, 2021, p. 536 ss.

[19] A tale riguardo, occorre evidenziare che l'art. 24, comma 3, l. n. 132/2025 delega il Governo ad adottare «uno o più decreti legislativi per adeguare e specificare la disciplina dei casi di realizzazione e di impiego illeciti di sistemi di intelligenza artificiale»; mentre il comma 5 dell'articolo in parola prevede, alla lett. d), che nell'esercizio di tale delega il Governo consideri, «nei casi di responsabilità civile, [la] previsione di strumenti di tutela del danneggiato, anche attraverso una specifica regolamentazione dei criteri di ripartizione dell'onere della prova, tenuto conto della classificazione dei sistemi di intelligenza artificiale e dei relativi obblighi come individuati» dall'*AI Act*. In argomento, v. le riflessioni di G. COMANDÈ, *La r.c. dopo la L. n. 132/2025: par favore parliamone!*, in *Danno resp.*, 2025, p. 673 ss.

[20] In termini generali, non va comunque dimenticato che dalla natura della responsabilità del medico discende, secondo i principi della materia, una diversa ricostruzione dell'onere della prova incombente sul paziente danneggiato (F. FERRETTI, *op. cit*., p. 60).

2. La responsabilità conseguente all'inadempimento degli obblighi generalmente imposti ai *deployer* dell'IA

L'affermazione della responsabilità del medico per i danni riportati dal paziente sottoposto ad una prestazione sanitaria eseguita con l'ausilio dell'IA può, sotto un primo punto di vista, conseguire all'inottemperanza degli obblighi generalmente imposti dalla legge agli utilizzatori dei sistemi intelligenti [21], sempre a condizione, s'intende, che si riscontri la sussistenza del nesso causale tra la violazione normativa commessa dal professionista sanitario e i pregiudizi lamentati dal malato [22].

In tal senso sembra che si possa concludere, innanzitutto, nelle ipotesi in cui a causare danno al paziente sia stato un utilizzo del sistema di IA inappropriato perché non conforme all'obbligo di attuare misure tecniche e organizzative atte a garantire il rispetto delle istruzioni che lo accompagnano (ai sensi dell'art. 13 *AI Act*) previsto dall'art. 26, par. 1, *AI Act*, senza peraltro dimenticare che, «in ragione dell'impiego di un dato sistema all'interno di un preciso contesto socio-tecnico, possono rendersi necessari accorgimenti tecnico-organizzativi aggiuntivi e complementari a quelli semplicemente correlati alle istruzioni ricevute dal *provider*» [23], la mancata o inadeguata adozione dei quali potrebbe pure costituire il fondamento per l'affermazione della responsabilità del medico.

La responsabilità in esame dovrebbe poter essere ravvisata, in secondo luogo, in capo al medico che abbia scorrettamente utilizzato un sistema di IA perché privo delle conoscenze, delle competenze e della formazione necessarie, tale comportamento contrastando con: l'esigenza di affidare la supervisione dei sistemi intelligenti a persone fisiche dotate della competenza, della formazione e dell'autorità necessarie, nonché di un adeguato sostegno, stabilito dall'art. 26,

[21] Nel senso che la violazione degli obblighi imposti dall'*AI Act* a fornitori e *deployer* di sistemi di IA può senz'altro fondare, pure in mancanza di un'espressa previsione normativa in tal senso, l'affermazione della responsabilità civile per i danni conseguentemente subiti da terzi, v., tra gli altri, A. GENTILI, *Regole per l'intelligenza artificiale*, in *Contr. impr.*, 2024, p. 1048; P. VALSECCHI, *Obblighi e responsabilità nella cornice normativa dell'AI Act. Commento agli artt. 16-26*, in M. D'AMICO-F. BALAGUER CALLEJÓN-A. AGUILAR CALAHORRO-P. GAMBATESA-S. DI GIOVANNI (a cura di), *op. cit.*, p. 353 ss.; N. MUCCIOLI, *La responsabilità sanitaria messa alla prova dall'intelligenza artificiale*, in *Riv. it. med. leg.*, 2025, p. 605.

[22] D. GIANTI, *op. cit.*, p. 204. In tema di nesso causale v. *infra*, *amplius*, § 6.

[23] A. MANTELERO, *Articolo 26 - Obblighi dei deployer dei sistemi di IA ad alto rischio*, in A. MANTELERO-G. RESTA-G.M. RICCIO (a cura di), *Intelligenza artificiale. Commentario*, Milano, 2025, p. 328.

par. 2, *AI Act* [24] in collegamento con il già menzionato art. 14 *AI Act* [25]; il parimenti già menzionato obbligo di alfabetizzazione in materia di IA contemplato dall'art. 4 *AI Act* [26]; il generale dovere di diligenza professionale nell'esercizio della professione medica [27], sulla scorta della quale il medico dovrebbe essere considerato allo stesso modo responsabile anche qualora abbia impiegato uno *smart device* per affrontare un caso clinico che invece richiedeva metodi di cura tradizionali [28] oppure si sia ingiustificatamente rifiutato di utilizzare un sistema

[24] Pare utile precisare che i termini «autorità» e «sostegno» menzionati dalla norma fanno riferimento alla necessità che la persona fisica deputata a supervisionare il sistema di IA abbia, rispettivamente, un'adeguata posizione gerarchica e funzionale all'interno dell'organizzazione del *deployer* e sufficienti risorse in termini di personale, mezzi tecnici e dotazioni economiche: M. LEFFI, *Articolo 26 - Obblighi dei deployer dei sistemi di IA ad alto rischio*, in P. PERRI-G. VACIAGO-G. ZICCARDI (a cura di), *op. cit*., p. 351.

[25] Come peraltro osserva A. MANTELERO, *Articolo 14 - Sorveglianza umana*, in A. MANTELERO-G. RESTA-G.M. RICCIO (a cura di), *op. cit*., p. 258, l'obbligo di progettare e sviluppare i sistemi di IA ad alto rischio in modo tale che possano essere efficacemente supervisionati dall'uomo durante l'utilizzo stabilito dalla norma in parola «pare irrealistico […], se interpretato in maniera letterale, salvo fare aggio sulla nozione di 'efficacia' per ammettere un livello di supervisione circostanziato e proporzionato alle reali possibilità del supervisore. […] Così, ad esempio, un oncologo potrà valutare in maniera completa ed efficace la diagnosi di cancro elaborata da un'IA di diagnostica per immagini prima di intervenire chirurgicamente sul paziente. Pur non conoscendo la ragione dell'eventuale errore della macchina (allucinazione, infra-rappresentazione di alcune categorie, ecc.) si potrà render conto dello stesso». Riflette approfonditamente sul punto anche U. RUFFOLO, *La responsabilità da scelte diagnostiche o terapeutiche suggerite dall'Intelligenza Artificiale generativa od operate da quella agentiva*, in C. DE MENECH-M. FOGLIA (a cura di), *Autodeterminazione, cura e ricerca nell'era digitale. Interessi antagonisti e modelli di bilanciamento*, Torino, 2026, p. 343 ss., il quale in particolare ritiene che «dovrebbero considerarsi sussistenti le caratteristiche di spiegabilità e trasparenza di una (anche opaca) IA quando, ancorché non immediatamente evidente, il suo livello di affidabilità e di rispetto dei valori e della dignità umani appare comunque suscettibile di adeguata verifica mediante sistemi e criteri che possono spaziare dalle più idonee modalità di testarne la funzionalità e la correttezza dell'operato alla controllabilità dei dati e parametri utilizzati nel *training* iniziale; processo, quest'ultimo, che appare solo eventuale».

[26] N. MUCCIOLI, *op. cit.*, p. 604 s.

[27] V., in particolare, A. ALBANESE, *op. cit.*, p. 314, il quale evidenzia che «affinché il professionista possa svolgere adeguatamente la propria attività è necessaria una formazione specifica all'utilizzo degli strumenti basati sull'IA. […] Ciò significa che il professionista non solo deve aggiornare le proprie competenze cliniche, ma deve anche conoscere i principi di funzionamento dei sistemi di IA, i loro limiti e le potenziali fonti di errore, le modalità corrette di interpretazione dei risultati forniti. In un contesto di sanità sempre più digitalizzata, l'aggiornamento continuo e la capacità di integrare correttamente l'IA nella pratica clinica rientrano tra gli obblighi professionali. Il possesso di una adeguata preparazione tecnica è un requisito intrinseco al concetto di perizia, il cui difetto integra un'ipotesi di colpa, criterio di imputazione della responsabilità […] del professionista».

[28] S. DE LUCIA, *op. cit.*, p. 1195; N. MUCCIOLI, *op. cit.*, p. 607; A. COLARUOTOLO, *op. cit.*, p. 284.

di IA che avrebbe garantito al paziente un trattamento più efficace [29].

Qualora si sia rivelata dannosa per il paziente, al medico dovrebbe potere essere parimenti rimproverata, in terzo luogo, l'inosservanza dell'obbligo di garantire che i dati di *input* dei sistemi di IA siano pertinenti e sufficientemente rappresentativi alla luce della finalità del sistema stesso previsto dall'art. 26, par. 4, *AI Act*, norma che viene del resto ripresa e integrata, sul versante del diritto nazionale, dall'art. 7, comma 6, l. n. 132/2025, ai sensi del quale i dati in discorso devono essere «affidabili, periodicamente verificati e aggiornati al fine di minimizzare il rischio di errori e migliorare la sicurezza dei pazienti» [30].

Da ultimo, la responsabilità del medico sembrerebbe potersi ricollegare pure al mancato adempimento dell'obbligo, previsto dall'art. 26, par. 5, *AI Act*, di monitorare il funzionamento del sistema di IA e sospenderne l'uso, informando i fornitori e le autorità pertinenti, qualora emerga che il suo utilizzo in conformità alle istruzioni ricevute [31] comporti rischi per la salute, la sicurezza o i diritti fondamentali delle persone o financo si verifichi un incidente grave [32]: comportando il permanere in attività di un dispositivo malfunzionante e caratterizzato da una spiccata pericolosità, difatti, una siffatta omissione potrebbe facilmente tradursi in un evento dannoso per la salute del paziente, che il medico dovrebbe invece essere in grado di tutelare anche eventualmente subentrando al sistema di IA che debba essere disattivato o cessi di operare autonomamente [33].

---

[29] F. FERRETTI, *op. cit.*, p. 171; E. CAMPAGNANO, *L'AI e la responsabilità medica aumentata: dalla colpa tradizionale al governo critico della tecnologia*, in *Riv. it. med. leg.*, 2025, p. 994.

[30] Come evidenzia A. MANTELERO, *Articolo 26 - Obblighi dei deployer dei sistemi di IA ad alto rischio*, cit., p. 328, il carattere della rappresentatività dei dati richiamato dal legislatore europeo «avrà certamente un ruolo importante con riguardo ai possibili *bias* derivanti da sotto-rappresentazione o sovra-rappresentazione di alcune categorie di soggetti, ma può anche essere intesa in senso più ampio, in termini di completezza degli *input* ai fini di fornire tutti gli elementi di cui il sistema necessita per un corretto funzionamento». Nel senso che «si potrebbe ipotizzare una responsabilità del medico per aver inserito nella macchina un pacchetto di dati incompatibile con la sua popolazione di pazienti, conducendo pertanto il mezzo a produrre un risultato falsato», v. pure E. COLLETTI, *op. cit.*, p. 244.

[31] Sebbene il dettato letterale della norma preveda che il *deployer* monitori il funzionamento del sistema (soltanto) sulla base delle istruzioni ricevute, è ragionevole pensare che in determinate situazioni il *deployer* debba spingersi oltre al fine di effettuare un monitoraggio del sistema adeguato e soddisfacente: A. MANTELERO, *op. loc. ult. cit.*; M. LEFFI, *op. cit.*, p. 353.

[32] Che l'art. 3, par. 1, n. 49 *AI Act* definisce come un incidente o malfunzionamento di un sistema di IA che, direttamente o indirettamente, causa una delle conseguenze seguenti: a) il decesso o gravi danni alla salute di una persona; b) una perturbazione grave e irreversibile della gestione o del funzionamento delle infrastrutture critiche; c) la violazione di obblighi stabiliti da norme del diritto dell'Unione e intesi a proteggere i diritti fondamentali; d) gravi danni alle cose o all'ambiente.

[33] G. SALITO, *op. cit.*, p. 861.

3. La responsabilità derivante dallo scorretto utilizzo degli *output* prodotti dall'IA in ambito sanitario

Rispetto alle vicende fin qui considerate, il giudizio sulla responsabilità del medico si fa decisamente più complesso nelle ipotesi in cui si tratta di valutare l'atteggiamento tenuto dal professionista sanitario rispetto alle indicazioni diagnostiche e/o terapeutiche fornite dall'IA.

Com'è stato icasticamente osservato, sotto questo punto di vista «gli scenari immaginabili sono intuitivi. […] Il medico può seguire una proposta diagnostico-terapeutica che con il senno di poi si rivela non corretta o non seguirne una che si rivela corretta […]. Sullo sfondo rimane il quesito di se, quando e quanto biasimo possa andare al professionista, al di là del fatto che formalmente la scelta diagnostico-terapeutica rimane fondamentalmente un suo atto medico» [34].

Al fine di rispondere al quesito in discorso, occorre ricordare che, come già accennato, il giudizio sulla responsabilità civile del medico è disciplinato dalla Legge Gelli-Bianco tramite due disposizioni: l'art. 5, che impone al professionista sanitario di attenersi, salve le specificità del caso concreto, alle raccomandazioni previste dalle linee guida validate e pubblicate dall'Istituto superiore di sanità nonché, in loro mancanza, alle c.d. buone pratiche clinico-assistenziali; l'art. 7, comma 3, alla stregua del quale il giudice, nella determinazione del risarcimento del danno, tiene conto della condotta tenuta dal medico ai sensi dell'articolo testé menzionato [35]. Come costantemente precisato dalla dottrina e

[34] G. COMANDÈ, *Tecnologie e metamorfosi del danno e della responsabilità sanitaria*, in *Danno resp*., 2024, p. 158. A queste ipotesi può essere, invero, affiancato l'ulteriore scenario in cui il medico, pur dando seguito alle indicazioni esatte fornite dalla macchina, cagioni un danno al paziente eseguendo in maniera scorretta un segmento della prestazione sanitaria – per esempio, la somministrazione della terapia consigliata dallo *smart device* – non intaccato dall'utilizzo dell'IA: in queste vicende, peraltro, il sistema di IA non gioca alcun ruolo nella causazione dell'evento lesivo e si configura, quindi, un caso "ordinario" di responsabilità sanitaria tradizionalmente intesa (F. FERRETTI, *op. cit.*, p. 169 s.).

[35] Secondo la prevalente dottrina, in questo modo la norma finisce per incidere non soltanto sull'*an*, ma pure sul *quantum* della responsabilità sanitaria, in particolare imponendo al giudice, oltre che di valutare se il sanitario abbia o meno rispettato le regole dell'*ars medica* espresse nelle linee guida e nelle buone pratiche al fine di decidere circa la sussistenza stessa della sua responsabilità, pure di diminuire l'entità del risarcimento dovuto dal professionista che, pur avendo commesso un qualche errore fonte di *malpractice*, abbia al contempo ottemperato l'obbligo di attenersi ai parametri di condotta previsti dall'art. 5 Legge Gelli-Bianco. In argomento v., *amplius*, G. D'AMICO, *Responsabilità civile degli esercenti le professioni sanitarie*, in *Enc. dir.*, *I tematici*, VII, *Responsabilità civile*, Milano, 2024, p. 806 ss.; M. FACCIOLI, *La quantificazione del danno nella responsabilità sanitaria secondo la Legge Gelli-Bianco (art. 7, commi 3 e 4, l. n. 24 del 2017)*, in *Contr. impr*., 2020, p. 1049 ss.; L. GUFFANTI PESENTI, *Il ruolo della condotta del medico*

dalla giurisprudenza, la salvezza delle specificità del caso concreto impone all'autorità giudicante di evitare ogni sbrigativo automatismo tra il riscontro dell'osservanza/violazione di linee guida (e buone pratiche clinico-assistenziali) da un lato e l'esclusione/affermazione della responsabilità del medico dall'altro: se giustificata dalle peculiarità del caso clinico di specie, l'inosservanza di quei parametri non può quindi fondare un addebito di colpa professionale, che può essere di converso rimproverata al medico che irragionevolmente segua le indicazioni di linee guida (e buone pratiche clinico-assistenziali) inadeguate rispetto alla concreta fattispecie da affrontare [36].

In un quadro normativo siffatto, il ruolo degli *output* prodotti dai sistemi di IA utilizzati in medicina sembra potere essere allora essere ricostruito nei termini che seguono.

In termini generali, si può innanzitutto affermare che, come per le linee guida, «anche per i sistemi di IA […] è ragionevole ritenere che la discrezionalità del giudice insita nel giudizio di colpa non possa subire alcuna limitazione, nel senso che egli resta libero di valutare se le circostanze concrete esigessero un approccio differente da quello suggerito dalla tecnologia in parola. Ne deriva che, alla mancata adesione alla proposta diagnostica e terapeutica dell'IA non conseguirà per il medico l'automatica rimproverabilità della propria condotta, dovendo accertarsi […] se tale scostamento, avuto riguardo alla condizione del paziente, fosse opportuno. Specularmente, la scelta del sanitario di allinearsi ai risultati del sistema di IA potrà comunque configurare gli estremi della condotta colposa, laddove le specificità del caso concreto avrebbero suggerito l'adozione di trattamenti […] differenti causalmente idonei ad evitare il danno poi in concreto patito dal paziente» [37].

---

nella quantificazione del risarcimento. Note sull'art. 7, co. 3, l. 8-3-2017 n. 24, in *Eur. dir. priv*., 2017, p. 1499 ss.; N.C. SACCONI, *Condotta dell'esercente la professione sanitaria e quantificazione del risarcimento*, in *Resp. civ. prev*., 2018, p. 1351 ss.; V. BACHELET, *Colpa medica e linee guida tra imperizia* in eligendo *e* in executivis, in *Nuova giur. civ. comm*., 2022, p. 648 s.

[36] Su questi aspetti, v. la dottrina citata nella nota precedente, cui *adde* F. CEREA, *op. cit.*, p. 177 ss. e F. FERRETTI, *op. cit.*, p. 142 ss.

[37] F. CEREA, *op. cit.*, p. 183. Nello stesso senso, v. F. FERRETTI, *op. cit.*, p. 146 s., la quale osserva pure che «sebbene la decisione della sola AI – prima della valutazione del medico – non possa dirsi ancora la "regola medica del caso concreto", essa ha comunque un contenuto intrinseco in parte già personalizzato, in misura maggiore rispetto alla linea guida astratta e generale. Il responso dell'AI è dato dalla combinazione tra il *software* "intelligente" e i dati sanitari del soggetto inseriti dal medico e rappresenta una "proposta di decisione" già ritagliata sul caso di specie. Questa precisazione non vanifica il parallelismo […] tra linea guida e decisione dell'AI, il minimo comune denominatore del quale è rappresentato dal ruolo decisivo attribuito al medico sulla scelta di seguire o no la proposta fornita, a prescindere dal grado di "personalizzazione" in essa contenuto».

Scendendo maggiormente nel dettaglio, il rilievo da attribuire agli *output* prodotti dall'IA sembra poter essere, poi, differentemente determinato avendo riguardo alle specifiche caratteristiche del sistema intelligente di volta in volta impiegato dal professionista sanitario.

Alcuni sistemi di IA diffusi nella pratica medica, chiamati *knowledge-based systems*, fondano la propria conoscenza su di un insieme predefinito di informazioni contenute in linee guida, buone pratiche, studi clinici e strumenti simili appositamente impiegati per programmarne il funzionamento. Costituendo il risultato della combinazione di conoscenze scientifiche consolidate con i dati riguardanti il paziente, le indicazioni provenienti dai sistemi in parola possono quindi essere a tutti gli effetti equiparate, alla stregua dell'art. 5 Legge Gelli-Bianco, a raccomandazioni previste dalle linee guida, se la base di conoscenza del sistema comprende linee guida validate e pubblicate dall'Istituto superiore di sanità, o diversamente, qualora le informazioni presenti nel sistema non siano state sottoposte a tale procedura di accreditamento, alle buone pratiche clinico-assistenziali [38].

Né l'una né l'altra di tali conclusioni può estendersi, invece, agli *output* prodotti dai sistemi di IA, detti *non-knowledge-based systems*, che grazie a tecniche di *machine learning* possono imparare autonomamente dalle proprie esperienze e/o individuare modelli nei dati clinici prescindendo da un insieme predefinito di conoscenze e di *input* dal quale apprendere oppure combinando quest'ultimo con dati e informazioni ulteriori: in queste ipotesi, in altri termini, la parificazione delle indicazioni fornite dal sistema a linee guida e buone pratiche clinico-assistenziali *ex* art. 5 Legge Gelli-Bianco potrebbe operare soltanto in seguito all'introduzione di un'apposita disposizione di legge [39].

Che si tratti di *knowledge-based systems* piuttosto che di *non-knowledge-based systems*, nell'assimilare, *de iure condito* o *de iure condendo*, le indicazioni fornite dall'IA a linee guida e buone pratiche clinico-assistenziali occorre comunque procedere con una certa prudenza e considerare che «l'utilizzo del sistema di IA potrebbe rappresentare una nuova frontiera della medicina difensiva,

[38] G. COMANDÈ, *op. ult. cit.*, p. 157 s.; F. FERRETTI, *op. cit.*, p. 152 ss.; A.G. GRASSO, *Diagnosi algoritmica errata e responsabilità medica*, in *Riv. dir. civ.*, 2023, p. 342 ss., il quale evidenzia come tale soluzione non incontri «il rischio che, qualificando come buone pratiche raccomandazioni provenienti da linee guida (o documenti equivalenti) non accreditate, si dia accesso a direttive non verificate dalle istituzioni nazionali competenti», perché le raccomandazioni in esame «devono essere state elaborate a seguito di un processo di revisione sistematica della letteratura e dei contributi scientifici nonché redatte da società scientifiche riconosciute. […] Sarà in ogni caso onere del medico dimostrare di avere esaminato criticamente la buona pratica all'atto di effettuare la scelta diagnostica o terapeutica».

[39] G. COMANDÈ, *op. ult. cit.*, p. 159; F. FERRETTI, *op. cit.*, p. 150; A.G. GRASSO, *op. cit.*, p. 347.

assi perniciosa, e capace di comprimere notevolmente la capacità di giudizio del professionista e la sua autonomia» [40], nella misura in cui potrebbe indurre il medico a recepire in maniera acritica gli *output* prodotti dall'IA per il timore di andare incontro a responsabilità qualora una scelta clinica difforme dal responso della macchina dovesse rivelarsi pregiudizievole per il paziente [41]. Né deve essere sottovalutato il rischio di *deskilling*, espressione che indica il progressivo impoverimento di competenze e capacità in possesso del professionista sanitario conseguente alla tendenza – sulla quale si tornerà tra breve – a riporre eccessivo affidamento (*over-reliance*) sulle indicazioni dell'IA nell'esercizio della propria attività [42].

Un ulteriore aspetto di rilievo, che non può essere approfondito nell'economia del presente lavoro, riguarda poi la scelta delle modalità operative attraverso le quali integrare il giudizio umano e gli *output* dell'IA, che debbono essere attentamente costruite, caso pero caso, alla luce delle caratteristiche concrete del sistema adoperato e del settore della medicina nel quale viene utilizzato, ponderando attentamente vantaggi e svantaggi di ogni soluzione astrattamente praticabile, che non potrà evidentemente non ripercuotersi sul giudizio di responsabilità del medico in caso di errore [43].

---

[40] A. BERTOLINI, *Dall'*imaging *ai sistemi esperti: la responsabilità del medico e le nuove frontiere della medicina difensiva*, in M. FACCIOLI (a cura di), *Profili giuridici dell'utilizzo della robotica e dell'intelligenza artificiale in medicina*, Napoli, 2022, p. 56.

[41] S. DE LUCIA, *op. cit*., p. 1204 s. Per una più ampia analisi sul punto, v. A. NATALE, *La medicina difensiva al tempo dell'intelligenza artificiale*, in *Bio LJ*, 2025, p. 371 ss., il quale osserva come la diffusione dell'IA potrebbe dare luogo anche ad ulteriori fenomeni di medicina difensiva, quali: l'esecuzione di esami supplementari al fine di dissipare i dubbi sull'affidabilità delle indicazioni fornite dai sistemi intelligenti; la predisposizione di abbondante documentazione per giustificare la scelta di aderire o meno all'*output* algoritmico; il ricorso a coperture assicurative specifiche per i danni cagionati utilizzando l'IA; il rifiuto di avvalersi di questa tecnologia per il timore eventuali contenziosi, a discapito dell'innalzamento della qualità delle cure che potrebbe comportare.

[42] V., *ex multis*, A. COLARUOTOLO, *op. cit.*, p. 278; D. GIANTI, *op. cit.*, p. 208 s.; E. PAVAN, *Intelligenza artificiale e modelli di responsabilità civile: riflessioni a margine dei recenti interventi normativi europei*, in *Danno resp*., 2025, p. 689; R. RANIERI, *I rischi dell'utilizzo dell'IA in campo medico-sanitario:* deskilling*,* over-reliance *e disumanizzazione del rapporto di cura*, in *Riv. it. med. leg.*, 2025, p. 961 ss.

[43] Con riguardo, per esempio, al settore della diagnostica per immagini, v. E. VERONA, *Errore diagnostico e intelligenza artificiale: profili di responsabilità nell'*imaging *medico*, in *Nuova giur. civ. comm.*, 2025, p. 1075 ss., la quale evidenzia che i sistemi utilizzati in tale ambito, detti CAD (*Computer Aided-Diagnosis*), possono essere impiegati in tre modi diversi (p. 1078), «a seconda del momento in cui i risultati del sistema vengono presentati all'operatore […]: una modalità "primo lettore", in cui il CAD viene impiegato per primo e il medico si limita a esaminare in sequenza le segnalazioni del sistema, riportando soltanto le lesioni che ritiene presenti;

Ad ogni modo, tra i diversi scenari di responsabilità che si possono presentare, ad attirare l'attenzione è soprattutto quello che vede il medico conformarsi ad indicazioni dell'IA scorrette in dipendenza di un malfunzionamento del sistema: da un lato e in termini generali, perché l'operatore umano, com'è già stato evidenziato, tende a recepire piuttosto che respingere l'*output* dei sistemi intelligenti [44]; dall'altro lato e con specifico riguardo al nostro ambito d'indagine, perché senz'altro si applica anche al settore dell'IA il principio, da tempo affermatosi nel campo della medicina tradizionale, che considera responsabile del danno subito dal paziente in conseguenza di un "incidente tecnologico" il medico che conosceva, o avrebbe potuto conoscere con una verifica improntata al criterio dell'ordinaria diligenza professionale, il *deficit* dei mezzi tecnici a propria disposizione [45].

Nell'ottica della responsabilità colposa, peraltro, si deve escludere che il medico possa essere chiamato a rispondere del malfunzionamento di un sistema di

una modalità "secondo lettore", in cui il medico compie una prima lettura non assistita dell'intero *dataset* e solo successivamente si avvale del sistema CAD per verificare se ci sono ulteriori anomalie che non sono state rilevate durante la prima verifica; una modalità "lettore concorrente", in cui si applica immediatamente lo strumento CAD, ma, a differenza del primo caso, il medico non valuta solo le lesioni segnalate dal sistema, ma dà un'interpretazione di tutti i dati disponibili». Nel caso in cui l'interazione uomo-macchina si risolva in un errore diagnostico, allora, sembra potersi affermare (p. 1082 s.) che, nell'ipotesi della "seconda lettura", ove «il radiologo interviene per primo, svolgendo una lettura autonoma dell'esame, ed usa il *software* per confermare o rivedere la sua decisione iniziale, […] si può ipotizzare che la sua condotta di adesione all'indicazione errata dell'IA sia censurabile soltanto qualora abbia commesso errori di valutazione evitabili da un operatore dotato di quella specializzazione». Estesa questa soluzione all'ipotesi della "lettura concorrente", in cui il radiologo vede le segnalazioni del sistema mentre sta refertando, ci si avvede del fatto che invece, nel caso della modalità di "prima lettura", «in cui viene data priorità all'interpretazione del sistema, con una revisione dei risultati da parte del medico soltanto successiva, il raggio d'azione del sanitario è già delimitato dalle indicazioni del dispositivo […]. Si può quindi supporre che laddove venga ammesso l'impiego in prima lettura del *software*, stante l'elevata affidabilità del sistema e il ruolo in seconda battuta del sanitario, il medico che abbia usato correttamente il dispositivo non risponderebbe delle omesse segnalazioni, la cui responsabilità ricadrebbe invece sul produttore che ha garantito la qualità del *software*».

[44] In argomento v., *ex multis*, A. MANTELERO, *Articolo 14 - Sorveglianza umana*, cit., p. 257 s. Di questa tendenza è consapevole, del resto, lo stesso legislatore europeo, che nell'art. 14, par. 4, lett. *b*), *AI Act* stabilisce che i sistemi di IA ad alto rischio devono essere progettati e sviluppati in modo tale che la persona fisica addetta alla loro supervisione possa «restare consapevole della possibile tendenza a fare automaticamente affidamento o a fare eccessivo affidamento sull'*output* prodotto da un sistema di IA ad alto rischio […], in particolare in relazione ai sistemi di IA ad alto rischio utilizzati per fornire informazioni o raccomandazioni per le decisioni che devono essere prese da persone fisiche».

[45] A. ALBANESE, *op. cit.*, p. 313; A. COLARUOTOLO, *op. cit.*, p. 284.

IA che fuoriesce dall'ambito del proprio controllo in virtù del suo elevato grado di autonomia, opacità e/o imprevedibilità [46], in queste ipotesi profilandosi piuttosto una responsabilità di quanti hanno partecipato alla realizzazione e alla fornitura dello *smart device* difettoso nonché, eventualmente, della struttura sanitaria per c.d. difetto di organizzazione [47]. Com'è stato condivisibilmente sottolineato, «particolarmente importanti a questo proposito sono le raccomandazioni del produttore sull'uso del sistema […]. Se le indicazioni del produttore non sono trasparenti, la decisione errata del medico dovrà essere considerata in modo meno rigoroso, ma se le istruzioni del produttore sono improntate alla chiarezza e richiedono un certo *standard* di preparazione dell'operatore sanitario o, ad esempio, indicano una certa percentuale di possibile errore diagnostico del dispositivo in relazione a particolari contesti o situazioni, la responsabilità del medico […] dovrà essere riconosciuta» [48].

4. L'incidenza dell'utilizzo dell'IA sull'applicazione dell'art. 2236 c.c. in ambito sanitario

Da un diverso ma complementare punto di vista, occorre pure chiedersi se l'utilizzo dell'IA nell'erogazione delle cure possa influire sull'applicazione della limitazione della responsabilità del professionista ai soli casi di dolo o colpa grave prevista dall'art. 2236 c.c. per il caso di prestazioni implicanti «la soluzione di problemi tecnici di speciale difficoltà» [49].

[46] F. CEREA, *op. cit.*, p. 189 s.; M. FACCIOLI, *Intelligenza artificiale*, cit., p. 736; s.; A. FIORENTINI, *op. cit.*, p. 1263; F. FERRETTI, *op. cit*., p. 170 ss.; N. MUCCIOLI, *op. cit.*, p. 609; S. DE LUCIA, *op. cit*., p. 1198 ss.; V. DE BERARDINIS, *L'impiego delle nuove tecnologie in medicina*, in G. ALPA (a cura di), *Diritto e intelligenza artificiale*, Pisa, 2020, p. 494; G. FIORINELLI-L. MENOZZI-A. SARACINO-M. SIMONI-A. PARZIALE, *Intelligenza artificiale in medicina e responsabilità: scenari giuridici e medico-legali*, in *Riv. it. med. leg.*, 2025, p. 840 s.; S. CACACE, *Le decisioni nella relazione di cura e le conseguenze dell'intelligenza artificiale*, in *Jus civ.*, 2025, p. 268.

[47] In argomento v., tra gli altri, F. CEREA, *op. cit.*, pp. 194 ss., 218 ss.; F. FERRETTI, *op. cit.*, p. 182 ss., 212 ss.; A. D'ADDA, *op. cit.*, p. 819 ss.; E. COLLETTI, *op. cit.*, pp. 31 ss., 231 ss.; A. ALBANESE, *op. cit.*, p. 309 ss.; M. FEOLA, *La responsabilità sanitaria e le nuove sfide dell'era digitale*, in R. FAVALE-M.A. MANTOVANI (a cura di), *op. cit*., p. 127 ss.

[48] A.G. GRASSO, *op. cit.*, p. 351.

[49] Come noto la norma, per quanto dettata in materia contrattuale, è dall'unanime giurisprudenza così come dalla maggioritaria dottrina ritenuta applicabile anche all'ambito della responsabilità aquiliana, in quanto esprime un criterio di giudizio che si ricollega alla natura stessa dell'attività regolamentata e prescinde dal contesto, contrattuale o meno, nel quale la medesima viene concretamente svolta: v., per tutti, C.M. BIANCA, *op. cit.*, p. 556 s.

Non essendovi dubbi circa la necessità di valutare il campo di applicazione della norma (anche) alla luce dello stato di avanzamento della scienza e della tecnica del momento storico nel quale si giudica la responsabilità in discorso [50], al quesito viene generalmente fornita una risposta positiva, accompagnata dalla precisazione che non sarebbe tuttavia corretto istituire un'automatica correlazione tra la sofisticatezza, l'innovatività, l'autonomia e l'opacità dei sistemi di IA impiegati in medicina e l'operatività dell'art. 2236 c.c.: secondo una consolidata impostazione, la ricorrenza del presupposto applicativo della norma dev'essere, infatti, sempre effettivamente riscontrata nel caso concreto, non essendo sufficiente, per dare spazio alla medesima, la potenziale prospettabilità di problemi di speciale difficoltà desunta da categorie astratte e predefinite [51].

Il ragionamento surriferito tende, peraltro, a considerare l'impiego dell'IA come un tendenziale indice della sussistenza della speciale difficoltà della prestazione contemplata dall'art. 2236 c.c., così dimenticando che i vantaggi in termini di accuratezza, precisione ed efficacia dell'atto medico apportati dall'utilizzo della tecnologia in esame potrebbero anche – se non più probabilmente – condurre all'esito opposto, ovverosia a concludere che una prestazione in passato considerata di speciale difficoltà non debba più essere reputata tale grazie all'ausilio dei sistemi intelligenti [52].

Non si può d'altro canto sottacere come la giurisprudenza sia, da sempre, estremamente riluttante a dare concreta applicazione all'art. 2236 c.c. – tanto in termini generali [53], quanto – in ambito sanitario [54] e ne circoscriva l'ambito di

[50] F. MARINELLI-F. CAROCCIA, *Contratto d'opera e prestazione d'opera intellettuale*, in *Trattato di diritto civile del Consiglio nazionale del notariato*, diretto da P. Perlingieri, Napoli, 2013, p. 271.

[51] G. SALITO, *op. cit.*, p. 862; A. NATALE, *op. cit*., p. 391; E. VERONA, *op. cit*., p. 1082; F. FERRETTI, *op. cit.*, p. 172; A.G. GRASSO, *op. cit.*, p. 351 s.; G. VOTANO, *Intelligenza artificiale in ambito sanitario: il problema della responsabilità civile*, in *Danno resp*., 2022, p. 675.

[52] F. CEREA, *op. cit.*, p. 185.

[53] M. GAZZARA, *"In difesa" dell'art. 2236 cod. civ.*, in *Nuovo dir. civ.*, 2020, p. 53; M. FRANZONI, *Dalla colpa grave alla responsabilità professionale*, IV ed., Torino, 2023, p. 60 s. Né va dimenticato che, sempre secondo la giurisprudenza, la limitazione di responsabilità prevista dell'art. 2236 c.c. va circoscritta all'aspetto della perizia del professionista, lasciando fuori dal suo ambito di riferimento i profili della diligenza e della prudenza: v., fra le tante, Cass., sez. III, 17 settembre 2024, n. 25026; Cass., sez. III, 15 febbraio 2022, n. 4905; Cass., sez. III, 10 marzo 2014, n. 5506; Cass., sez. III, 27 ottobre 2011, n. 22398; Cass., sez. III, 13 gennaio 2005, n. 583.

[54] Nell'area della responsabilità sanitaria, la giurisprudenza ritiene che il caso affidato al medico possa considerarsi di particolare difficoltà solo quando esso richieda un impegno in-

operatività entro confini che in dottrina sono prevalentemente e a ragione considerati eccessivamente restrittivi [55]: prima di delimitarne ulteriormente la sfera di operatività facendo leva sull'utilizzo dell'IA, quindi, occorrerebbe restituire alla norma un ruolo adeguato alla sua *ratio*, che come si legge nella *Relazione al codice*, n. 917, è quella di conciliare «due opposte esigenze, quella di non mortificare l'iniziativa del professionista col timore di ingiustificate rappresaglie da parte del cliente in caso di insuccesso e quella inversa di non indulgere verso non ponderate decisioni o riprovevoli inerzie del professionista» stesso [56].

tellettuale e una preparazione professionale superiori a quelli del professionista medio, il che dovrebbe intendersi accadere, precisamente, soltanto quando il caso stesso si presenti come eccezionale e straordinario, per non essere ancora stato adeguatamente studiato nella scienza e sperimentato nella pratica, ovvero per essere oggetto di dibattiti scientifici con sperimentazione di sistemi diagnostici, terapeutici e di tecnica chirurgica diversi o incompatibili (così, *ex multis*, già Cass., sez. III, 26 marzo 1990, n. 2428; ma per affermazioni del tutto simili riferite alla responsabilità dell'ingegnere, v. Cass., sez. III, 31 luglio 2015, n. 16275). Con puntuali interventi la Suprema Corte ha, inoltre, ulteriormente precisato che: la difficoltà dell'intervento non si identifica con l'aleatorietà, perché anche l'effetto di una prestazione terapeutica di facilissima esecuzione, come un'iniezione o la somministrazione di un farmaco, può in talune ipotesi apparire del tutto aleatorio (Cass., sez. III, 21 giugno 2004, n. 11488); anche quando ci si trovi di fronte ad un intervento chirurgico che presenta il carattere della speciale difficoltà, la limitazione di responsabilità prevista dall'art. 2236 c.c. non si estende alla fase post-operatoria, nella quale il medico risponde, anche per colpa lieve, dell'inadempimento del dovere di vigilare attentamente sul paziente e adottare tutte le precauzioni e i rimedi conosciuti dalla scienza e dalla pratica medico-specialistica al fine di evitare che insorgano complicazioni (Cass., sez. III, 28 settembre 2009, n. 20790); il requisito della speciale difficoltà della prestazione non è integrato dall'incompletezza o dalla reticenza delle informazioni fornite dal paziente circa le proprie condizioni di salute, laddove queste ultime siano accertabili attraverso la corretta esecuzione della prestazione curativa (Cass., sez. III, 12 settembre 2013, n. 20904); ancora, un intervento chirurgico di norma routinario non può mai ritenersi di speciale difficoltà per il solo fatto che nel corso di esso si verifichino delle complicanze (Cass., sez. III, 22 novembre 2012, n. 20586).

[55] Tra gli altri v., anche per ulteriori citazioni in tal senso, M. GAZZARA, *op. cit.*, spec. p. 72 ss.; C. GRANELLI, *La medicina difensiva in Italia*, in *Resp. civ. prev.*, 2016, p. 32; M. ROSSETTI, Unicuique suum, *ovvero le regole di responsabilità non sono uguali per tutti (preoccupate considerazioni sull'inarrestabile fuga in avanti della responsabilità medica)*, in *Giust. civ.*, 2010, I, p. 2224 ss.; M. FOGLIA, *Il prisma della prestazione medica e l'"azzeramento" dell'art. 2236 cod. civ.*, in *Nuova giur. civ. comm.*, 2010, p. 372 ss.

[56] Cfr. F. CEREA, *op. cit.*, p. 185, secondo cui «l'applicazione della disposizione entro gli angusti confini che fino ad oggi l'hanno caratterizzata non lasciano, dunque, presagire una sua futura più ampia operatività a seguito dell'ingresso dell'IA nella prestazione medica».

5. L'inconfigurabilità di una responsabilità oggettiva del medico per i danni cagionati al paziente impiegando l'IA

Le riflessioni e le conclusioni fino a qui elaborate potrebbero essere, invero, messe in discussione attraverso diversi percorsi ermeneutici che, in un modo o nell'altro, finirebbero per incentrare il giudizio sulla responsabilità del medico utilizzatore di sistemi di IA su di un criterio di imputazione di natura oggettiva (o semi-oggettiva) anziché sul criterio della colpa, con la conseguenza di estendere i contorni e intensificare il rigore della responsabilità di cui si discute.

A suggerire questa direzione potrebbe essere, innanzitutto, la riqualificazione dell'attività sanitaria eseguita con l'ausilio dell'IA da prestazione di mezzi a prestazione di risultato, con tutto ciò che notoriamente ne consegue in punto di oggettivizzazione della responsabilità del professionista, operata sulla scorta di quegli orientamenti giurisprudenziali che già individuano diverse ipotesi in cui ritengono sussistere gli estremi per ricostruire un'obbligazione di risultato in capo al medico o, comunque, fanno uso di regole operazionali che finiscono per produrre esiti applicativi del tutto simili [57]. Tuttavia, per quanto l'IA possa incrementare l'accuratezza, la precisione e l'efficacia dell'atto medico, il suo utilizzo non sembra *sic et simpliciter* imporre lo sviluppo di un siffatto ragionamento, avulso dalle circostanze del caso concreto: l'impiego di macchine intelligenti sembra infatti costituire un fattore neutrale rispetto all'individuazione della natura della prestazione sanitaria, che in assenza di ulteriori (rispetto al coinvolgimento dell'IA) elementi idonei a ricondurla nell'ambito delle prestazioni di risultato dovrebbe normalmente rientrare nell'ambito delle prestazioni di mezzi, la responsabilità per l'inadempimento delle quali è per l'appunto dominata dal criterio della colpa professionale [58].

[57] Com'è stato evidenziato in dottrina, questi orientamenti si fondano sull'osservazione del fatto che l'esecuzione di determinati trattamenti sanitari è in certi casi presieduta da regole tecniche molto specifiche e altamente vincolanti, al punto da potersi istituire, anche con il conforto del dato statistico, uno stretto collegamento tra il rispetto di quelle regole e il raggiungimento di un certo esito clinico e, di conseguenza, tra il fallimento delle cure da un lato e la sussistenza di un errore nell'esecuzione del trattamento dall'altro. In argomento v., tra gli altri, M. FACCIOLI, *Mezzi e risultato nelle prestazioni sanitarie*, in *Resp. med.*, 2023, p. 137 ss.; G. D'AMICO, *Responsabilità per inadempimento e distinzione tra obbligazioni di mezzi e di risultato*, in *Il diritto delle obbligazioni e dei contratti: verso una riforma? Atti del Convegno per il cinquantenario della Rivista di diritto civile*, Padova, 2006, p. 154 ss.; G. SICCHIERO, *Dell'adempimento. Artt. 1176-1179*, in *Il codice civile. Commentario*, fondato da P. Schlesinger e diretto da F.D. Busnelli, Milano, 2016, p. 194 ss.; A. DI MAJO, *La salute responsabile*, Torino, 2018, p. 29 s.

[58] M. FACCIOLI, *Intelligenza artificiale*, cit., p. 736; A. FIORENTINI, *op. cit*., p. 1261; G. VOTANO, *op. cit.*, p. 674 s.; V. DI GREGORIO, *Robotica e intelligenza artificiale: profili di r.c. in*

Una soluzione di segno contrario rispetto a quella fin qui accolta è stata, poi, da taluno sostenuta facendo leva sull'istituto della responsabilità indiretta e oggettiva del debitore per il fatto degli ausiliari contemplata dall'art. 1228 c.c., ritenuto applicabile in via analogica alle vicende che vedono coinvolti il medico libero professionista obbligato all'esecuzione della prestazione sanitaria e lo strumento di adempimento costituito dal sistema di IA [59]. In questa prospettiva, si è affermato che «l'utilità tratta […] dal medico libero professionista mediante l'utilizzo della macchina autonoma sembra veicolare la responsabilità verso la sfera di rischio dell'utilizzatore, anche quando il sinistro derivi da scelte autonome, imprevedibili ed opache, della macchina. Sicché, anche quando non sia accertata un'imperizia nell'esecuzione della prestazione (che sarebbe ragione *ex se* di responsabilità per colpa) non dovrebbe essere preclusa l'operatività di criteri di responsabilità più obiettivi: nella stessa logica per cui il debitore è responsabile del fatto dell'ausiliario» [60]. Per quanto suggestiva, tale proposta ermeneutica non può essere tuttavia condivisa per due ordini di ragioni: innanzitutto, per la ben nota irriferibilità ai sistemi di IA dell'attributo della soggettività giuridica che tale impostazione necessariamente finisce per presupporre [61];

campo sanitario, in *Resp. med.*, 2019, p. 442; E. COLLETTI, *op. cit.*, p. 230 s.; S. DE LUCIA, *op. cit*., p. 1195; A. NATALE, *op. cit*., p. 43; F. FERRETTI, *op. cit*., p. 64 ss., la quale sviluppa tali considerazioni con riguardo alle prestazioni sanitarie eseguite con l'ausilio della robotica.

[59] A. D'ADDA, *op. cit.*, p. 813 ss., il quale ritiene potersi applicare tale ragionamento tanto alla struttura sanitaria quanto al medico libero professionista osservando, con riguardo a quest'ultimo, che «la sua scelta di servirsi non di strumenti convenzionali, bensì di strumenti autonomi, è analoga a quella operata da chi si avvale di ausiliari; ed il rischio di siffatta autonomia non può che ricadere sul debitore, le cui scelte esecutive restano a lui integralmente imputabili».

[60] A. D'ADDA, *op. cit.*, p. 815. Pur senza richiamare l'art. 1228 c.c., giungono sostanzialmente alle stesse conclusioni G. DI ROSA, *op. cit.*, p. 227 e F. FERRETTI, *op. cit.*, p. 185, la quale in particolare ritiene che «il medico libero professionista […] assume su di sé non solo le responsabilità proprie dell'attività clinica – che trovano un limite nella diligenza esigibile – ma anche quelle organizzative e imprenditoriali, tra cui rientra anche la responsabilità per il comportamento imprevedibile di sistemi di AI […]. In tale prospettiva, la scelta del medico di avvalersi di dispositivi intelligenti colloca il malfunzionamento dell'apparato nell'alveo del rischio professionale assunto consapevolmente, non solo attraverso la stipula del contratto con il paziente, ma anche e soprattutto per effetto dell'organizzazione autonoma dei mezzi di cura».

[61] Irriferibilità sulla scorta della quale, come pure è ormai diffusamente riconosciuto, si esclude la praticabilità di una soluzione che riconoscesse i sistemi intelligenti direttamente responsabili dei danni cagionati a terzi: in argomento v., *ex plurimis*, E. MINERVINI, *Note brevi sull'intelligenza artificiale*, in questa *Rivista*, 2025, p. 193, secondo cui «è [...] un non problema quello della soggettività della macchina intelligente o se si preferisce della personalità elettronica: attribuire tale personalità equivarrebbe a nominare senatore il cavallo, come ebbe a fare Caligola, abbandonandosi a fantasiose discussioni»; M.F. CURSI, *IA: modelli tradizionali di re-*

in secondo luogo, per l'impossibilità di rimproverare alla macchina l'elemento soggettivo del dolo o della colpa richiesto dall'art. 1228 c.c. perché il debitore sia chiamato a rispondere per i danni cagionati dai suoi ausiliari [62].

Sostanzialmente per le stesse ragioni, del resto, si deve escludere che, nella prospettiva della responsabilità extracontrattuale, il medico possa essere chiamato a rispondere dei danni cagionati dai sistemi di IA utilizzati per le cure sulla scorta dell'applicazione analogica dell'art. 2049 c.c. in tema di responsabilità dei padroni e dei committenti per il fatto illecito dei loro domestici e commessi [63].

Poco convincente è pure la tesi che, sempre inquadrando il tema nell'ambito aquiliano, vorrebbe ricondurre la responsabilità per i danni cagionati dalle prestazioni sanitarie eseguite tramite l'IA sotto il regime dell'art. 2050 c.c. [64], anche alla luce dell'approccio incentrato sulla valorizzazione del profilo del rischio caratterizzante questo tipo di tecnologia adottato dal legislatore europeo nell'*AI Act*: così sollevato il paziente danneggiato dall'onere di provare la colpevolezza del convenuto, la prova liberatoria prevista dalla norma *de qua* in capo a quest'ultimo verrebbe allora identificata, alla luce della rapidità dell'evoluzione tecnologica in materia, nell'adozione non solo delle prescrizioni tecniche

sponsabilità e nuove prospettive, in questa *Rivista*, 2026, p. 12, la quale evidenzia che «l'innovativa proposta non risolve il problema di chi sia tenuto a pagare in caso di danno prodotto dalla macchina. Oltre al fatto che, attribuendo un'autonoma personalità alla macchina intelligente, si potrebbe correre il rischio di deresponsabilizzare chi poteva essere chiamato a rispondere vicariamene, con l'effetto di vedere limitato il risarcimento al fondo patrimoniale attribuito alla macchina intelligente».

[62] V., tra gli altri, F. CEREA, *op. cit.*, p. 212 s.; F. FERRETTI, *op. cit.*, p. 265 s.; E. PAVAN, *op. cit.*, p. 691; A. GENTILI, *op. cit.*, p. 1062 s.; G. PROIETTI, *La responsabilità nell'intelligenza artificiale e nella robotica. Attuali e futuri scenari nella politica del diritto e nella responsabilità contrattuale*, Milano, 2020, p. 177; A. ASTONE, *Sistemi intelligenti e regole di responsabilità*, in *Pers. merc.*, 2023, p. 496.

[63] V., tra gli altri, G. SALITO, *op. cit.*, p. 853 s.; F. LAGIOIA, *op. cit.*, p. 37; A. GENTILI, *op. cit.*, p. 1064 s.; M. COSTANZA, *L'Intelligenza artificiale e gli stilemi della responsabilità civile*, in *Giur. it.*, 2019, p. 1688 s.; C. SCOGNAMIGLIO, *Responsabilità civile ed intelligenza artificiale: quali soluzioni per quali problemi?*, in *Resp. civ. prev.*, 2023, p. 1086 s.; T. DE MARI CASARETO DAL VERME, *Intelligenza artificiale e responsabilità civile. Uno studio sui criteri di imputazione*, Napoli, 2024, p. 145 ss.

[64] Oltre alla dottrina citata alla nota 66, in questo senso v., per esempio, E. COLLETTI, *op. cit.*, p. 26 ss.; U. RUFFOLO, *L'Intelligenza Artificiale nei dispositivi medici e nell'attività sanitaria: per un'analisi interdisciplinare circa responsabilità, controlli preventivi e disciplina dei dati sanitari. Le conseguenti proposte interpretative e normative*, in U. RUFFOLO-M. GABBRIELLI (a cura di), *Intelligenza Artificiale, dispositivi medici e diritto. Un dialogo fra saperi: giuristi, medici e informatici a confronto*, Torino, 2023, p. 16 ss.; C. LEANZA, *Intelligenza artificiale e diritto: ipotesi di responsabilità civile nel terzo millennio*, in *Resp. civ. prev.*, 2021, p. 1023.

previste dal summenzionato regolamento europeo [65], ma pure delle misure di sicurezza ritenute idonee secondo il progresso scientifico del momento storico anche se non ancora recepite dalla normativa di settore [66]. Già ampiamente controversa in termini generali, la qualificazione dell'attività medica quale attività pericolosa ai sensi della norma testé richiamata risulta essere, infatti, del tutto implausibile quando la stessa si accompagna all'impiego di soluzioni tecnologiche avanzate che, lungi dall'introdurre fattori di pericolo, dovrebbero accrescere il livello di affidabilità, precisione e sicurezza delle attività svolte impiegando questa tecnologia, riducendo i rischi riconducibili all'errore umano [67]. Sul più generale piano della politica del diritto, inoltre, sembra sconsigliare il ricorso all'art. 2050 c.c. pure la considerazione della necessità di non disincentivare la produzione e la diffusione dei sistemi di IA tramite l'applicazione di un regime di responsabilità rigoroso nei confronti del danneggiante come quello delineato dalla norma in parola [68].

Maggiormente attendibile appare, almeno a prima vista, la tesi che ritiene applicabile alle fattispecie in esame la disciplina dell'art. 2051 c.c. sul danno da cose in custodia, interpretata in senso evolutivo e in collegamento sistematico con l'art. 2052 c.c., il quale invero assoggetta alla medesima prova liberatoria del caso fortuito la responsabilità per danni cagionati da animali, cose "animate" e dinamiche proprio come le macchine intelligenti [69]. Pure questa proposta ermeneutica è stata, tuttavia, efficacemente respinta mettendo in evidenza la dif-

[65] Il mancato rispetto delle quali dovrebbe essere peraltro sufficiente a decretare il mancato raggiungimento della prova liberatoria prevista dall'art. 2050 c.c.

[66] A. ASTONE, *op. cit.*, p. 500; E. FAZIO, *Intelligenza artificiale e diritti della persona*, Napoli, 2023, p. 214 s.; A. AMIDEI, *La proposta di Regolamento UE per un* Artificial Intelligence Act*: prime riflessioni sulle ricadute in tema di responsabilità da Intelligenza Artificiale*, in *Tecn. dir*., 2022, p. 16 ss.

[67] M. COSTANZA, *op. cit.*, p. 1688; A. NATALE, *op. cit.*, p. 392 s.; C. PERLINGIERI, *op. cit.*, p. 923 s.; V. DI GREGORIO, *op. cit.*, p. 444; G. SALITO, *op. cit*., p. 854; M. FEOLA, *op. cit*., p. 124; T. DE MARI CASARETO DAL VERME, *op. cit.*, p. 150 ss.

[68] Riflette approfonditamente sul punto M. GRONDONA, *L'art. 2050 nella prospettiva dell'IA medica*, in *Resp. med.*, 2025, p. 455 ss.

[69] V., tra gli altri, A. ASTONE, *op. cit.*, p. 497 ss.; T. DE MARI CASARETO DAL VERME, *op. cit.*, p. 288 ss.; M. RATTI, *Riflessioni in materia di responsabilità civile e danno cagionato da dispositivo intelligente alla luce dell'attuale scenario normativo*, in *Contr. impr*., 2020, p. 1182; L. DI DONNA, *Intelligenza artificiale e rimedi risarcitori*, Padova, 2022, p. 137 ss.; G. D'ALFONSO, *Il regime di responsabilità da cose in custodia tra questioni tradizionali e "responsabilità da algoritmo"*, in *EJPLT*, 2022, p. 101 ss., secondo cui l'applicabilità dell'art. 2051 c.c. andrebbe verificata distinguendo a seconda del controllo esercitabile sulla macchina e sulla facoltà dell'utilizzatore del sistema di IA di modificare le condizioni di rischio accedendo a servizi e/o prodotti *software* offerti da soggetti diversi dal realizzatore del sistema.

ficoltà di conciliare il criterio della custodia contemplato dall'art. 2051 c.c. con l'autonomia e l'imprevedibilità che caratterizza i sistemi di IA [70]: «la cosa, nel disegno dell'art. 2051 c.c., è entità lontana dalla A.I.» e «anche una interpretazione in senso estensivo della norma […] mantiene intatta la condizione del collegamento fra l'evento dannoso e le dinamiche dello stato o della condizione della cosa. La prospettiva rimane all'interno delle logiche della custodia. Il livello tecnologico della A.I. ed il suo contenuto standard, funzione/elaborazione, non aderisce alla realtà ripresa nel Codice Civile, ancora legato alla nozione di forza, anche se incrementata da macchine. L'intelligenza artificiale "provvederebbe" alla macchina, sicché di ciò che proviene da questa sarebbe fattore determinante. Il mondo delle cose senza responsabilità è troppo semplice di fronte alla intelligenza artificiale ed alle sue alchimie mentali» [71]. Occorre sottolineare, inoltre, che l'art. 2051 c.c. non potrebbe, in ogni caso, trovare spazio con riguardo al medico dipendente, la figura professionale di gran lunga più diffusa nel mondo della sanità, in forza del consolidato orientamento che considera la norma *de qua* inapplicabile a colui che, in qualità di lavoratore subordinato, detiene la cosa sotto la direzione e il controllo di altri, che ne ha la disponibilità giuridica e ne determina le condizioni di utilizzo e conservazione [72].

Contro l'applicabilità delle forme speciali di responsabilità previste negli artt. 2048 ss. c.c. alle fattispecie di nostro interesse può essere, infine, evidenziato che l'art. 7, comma 3, Legge Gelli-Bianco sottopone esplicitamente proprio la responsabilità della figura professionale da ultima richiamata alla disciplina della responsabilità per colpa tratteggiata dall'art. 2043 c.c. Ciò sembra precludere operazioni ermeneutiche dirette ad assoggettare la responsabilità in esame a diverse e più rigorose previsioni normative: la norma *de qua* è stata infatti introdotta allo scopo di contrastare il già menzionato fenomeno della medicina difensiva alleggerendo la pressione giudiziaria gravante sul personale medico e incanalando le azioni risarcitorie dei pazienti nei confronti delle strutture sanitarie [73], finalità che non sembra possano essere allora ostacolate

[70] Questo aspetto non sfugge ad A. ASTONE, *op. cit*., p. 499, il quale lo ritiene tuttavia superabile osservando che «la circostanza che il proprietario o chi utilizza il dispositivo autonomo non abbia di fatto un potere di controllo su di esso, idoneo a neutralizzarne le potenzialità dannose, non costituisce elemento costitutivo della fattispecie di responsabilità in questione né elemento di cui tener conto per escludere l'obbligazione risarcitoria in capo al custode».

[71] Sono parole di M. COSTANZA, *op. cit.*, p. 1687 s. Per analoghe considerazioni, v. pure C. SCOGNAMIGLIO, *op. cit*., p. 1088; V. DI GREGORIO, *op. cit.*, p. 58; G. SALITO, *op. cit.*, p. 854.

[72] A. D'ADDA, *op. cit.*, p. 816; A. ALBANESE, *op. cit*., p. 312.

[73] In argomento v., *ex multis*, C. GRANELLI, *Il fenomeno della medicina difensiva e la legge di riforma della responsabilità sanitaria*, in *Resp. civ. prev.*, 2018, p. 410 ss.; E. MARCHISIO,

dall'interprete irrigidendo in via ermeneutica il regime della responsabilità del medico ausiliario tratteggiato dal dettato normativo [74].

6. L'incidenza dell'utilizzo dell'IA sulla ricostruzione del nesso causale intercorrente tra la condotta del medico e il danno subito dal paziente

La pretesa risarcitoria dispiegata dal paziente per ottenere il risarcimento del danno cagionato dalle cure erogate con l'ausilio dell'IA deve infine confrontarsi, alla pari di tutte le azioni di responsabilità da *medical malpractice*, con le questioni relative al nesso eziologico, che nell'ambito in esame sembra dovere essere senz'altro concepito, in conformità ai principi generali della materia, secondo la nota impostazione giurisprudenziale incentrata sul criterio del "più probabile che non" [75].

Com'è stato da più parti evidenziato, la ricostruzione del nesso causale potrebbe peraltro presentarsi particolarmente complessa nella materia che ci occupa, perché a fronte dell'opacità, dell'autonomia e dell'imprevedibilità dei sistemi di IA è plausibile che possa non di rado emergere più di qualche difficoltà nello stabilire se il danno subito dal paziente debba essere o meno ricondotto ad un difetto di funzionamento della macchina, ad un errore umano nell'utilizzo del sistema oppure ancora al naturale peggioramento delle condizioni di salute del malato [76].

Secondo taluno, le surriferite caratteristiche dell'IA sarebbero anzi tali da interrompere il nesso di causalità tra l'operato del medico e il danno subito dal paziente, privando di ogni rilievo l'indagine sull'elemento soggettivo della condotta

Evoluzione della responsabilità civile medica e medicina "difensiva", in *Riv. dir. civ.*, 2020, p. 189 ss. Sulla possibilità che l'utilizzo di soluzioni tecnologiche particolarmente avanzate incrementi le aspettative dei pazienti e finisca per esasperare il problema della medicina difensiva, v., poi, M. CASALI-U. GENOVESE, *Intelligenza artificiale e responsabilità sanitaria: il punto di vista del medico legale*, in M. FACCIOLI (a cura di), *op. cit*., p. 117 ss.

[74] M. FACCIOLI, *Intelligenza artificiale*, cit., p. 738; N. MUCCIOLI, *op. cit*., p. 606 s.; E. PAVAN, *op. cit*., p. 691. In termini più generali, si esprime contro l'aggravamento, rispetto alle regole ordinarie della *medical malpractice*, della responsabilità del personale medico che utilizza l'IA anche M. TAMPIERI, *L'intelligenza artificiale e le responsabilità nel settore della medicina*, in *Resp. med*., 2024, p. 37.

[75] F. CEREA, *op. cit*., p. 186; A. COLARUOTOLO, *op. cit.*, p. 286.

[76] Cfr., tra gli altri, F. LAGIOIA, *op. cit.*, p. 109; G. VOTANO, *op. cit.*, p. 678; R. SCOTTI, *op. cit.*, p. 175 ss.; A. COLARUOTOLO, *op. cit.*, p. 286 s.; E. CAMPAGNANO, *op. cit.*, p. 988 ss.; S. MAULONI-M. MAZZANTI-L. BUSCEMI, *La nuova era dell'intelligenza artificiale: problemi medico-legali*, in *Riv. it. med. leg*., 2022, p. 698 s.

tenuta dal professionista sanitario [77]. Tale conclusione non può essere, tuttavia, in alcun modo condivisa, in quanto permetterebbe di andare esente da responsabilità, invocando l'autonomia, l'opacità e l'imprevedibilità dell'IA, al medico che abbia fatto acriticamente affidamento sulle indicazioni del sistema in contrasto con i doveri di controllo e supervisione sulle macchine intelligenti che sono stati esaminati nelle pagine precedenti [78].

Occorre piuttosto osservare che la questione in esame non è affatto inedita nella materia della responsabilità sanitaria, ove la difficoltà di ricostruire il nesso eziologico già da tempo emerge con una certa frequenza in virtù di fattori che sono essenzialmente ricollegati all'imprevedibilità delle reazioni dell'organismo umano alle cure mediche e che sotto questo punto di vista prescindono dalla circostanza che si faccia o meno ricorso all'IA [79].

Il problema probatorio della c.d. causa ignota costituisce, infatti, uno dei profili della responsabilità sanitaria maggiormente approfonditi tanto in sede dottrinale quanto nelle aule di giustizia negli ultimi anni, venendo da ultimo risolto con il noto indirizzo giurisprudenziale secondo cui il paziente è tenuto a provare, anche attraverso presunzioni, il nesso di causalità materiale tra la condotta del medico e l'evento dannoso, consistente nella lesione della salute o nell'aggravamento della situazione patologica o nell'insorgenza di una nuova malattia, mentre è onere della controparte, ove il paziente abbia dimostrato il suddetto nesso di causalità materiale, provare – di avere esattamente adempiuto, oppure – che l'inadempimento della prestazione sanitaria è dipeso da impossibilità della

[77] A. NATALE, *op. cit.*, p. 397; M. SAVINI NICCI, *Intelligenza Artificiale e responsabilità civile di strutture e professionisti sanitari*, in F. TRUBIANI (a cura di), *Sistemi di intelligenza artificiale in medicina verso lo spazio europeo dei dati sanitari. Un dialogo multidisciplinare*, Torino, 2025, p. 70.

[78] F. CEREA, *op. cit*., p. 187 s. Si soffermano sulla tesi esposta nel testo anche G. FIORINELLI-L. MENOZZI-A. SARACINO-M. SIMONI-A. PARZIALE, *op. cit.*, p. 846 s., per rilevare che «una siffatta impostazione – per quanto sicuramente valida in alcuni limitati casi […] – deve, tuttavia, essere riconciliata con le […] considerazioni di carattere tecnico, che riconducono la (relativa) controllabilità dell'*output* dell'IA alla qualità del *dataset*, al processo di *testing* e alla verifica dell'accuratezza: non può trascurarsi, dunque, come la stessa "imprevedibilità" del sistema risulti un rischio, nel complesso suscettibile di gestione e controllo da parte degli agenti "umani", rispetto ai quali può arretrare l'accertamento del nesso di causalità. Del resto, non va trascurato come l'AI Act annoveri anche l'*explainability* e la *traceability* tra i requisiti di sicurezza dei sistemi di IA ad alto rischio […]: sì che, oltre a facilitare *ex post* la "tracciabilità tecnica" del processo che ha determinato l'*output*, tali requisiti nuovamente ricollegano alla decisione e al controllo umani il gradiente di imprevedibilità del sistema di IA».

[79] Cfr. J.-S. BORGHETTI, *How can Artificial Intelligence be defective*, in S. LOHSSE-R. SCHULZE-D. STAUDENMAYER (a cura di), *Liability for artificial intelligence and the Internet of things*, Baden-Baden, 2019, p. 75.

prestazione derivante da causa non imputabile. Pertanto, se rimane ignota la causa del danno riportato dal paziente, sarà quest'ultimo a subirne le conseguenze vedendo la propria azione respinta, mentre se rimane ignota la causa che ha impedito ai medici e al nosocomio la corretta esecuzione della prestazione sanitaria, saranno questi ultimi a farne le spese tramite la condanna al risarcimento [80].

Le incertezze sulla ricostruzione del nesso eziologico che dovessero emergere nelle controversie per danni cagionati da trattamenti sanitari eseguiti con l'ausilio dell'IA dovrebbero potere essere, quindi, affrontate e risolte tramite l'applicazione di questi principi [81], nell'ambito dei quali un importante ruolo nel superamento delle difficoltà probatorie incontrate dal paziente potrebbe essere svolto dalla possibilità, come già detto ammessa dalla giurisprudenza, di avvalersi del meccanismo presuntivo [82].

[80] In argomento v., per tutti, N. RIZZO, *La causalità civile*, Torino, 2022, p. 72 ss.

[81] M. FACCIOLI, *op. cit*., p. 739 s.; A. COLARUOTOLO, *op. cit*., p. 287; N. MUCCIOLI, *op. cit*., p. 613 s. Assume una posizione critica al riguardo R. SCOTTI, *op. cit.*, p. 177, secondo cui «trasponendo (*mutatis mutandis*) tali principi, dettati dalla giurisprudenza più recente con riferimento a vicende di responsabilità (soltanto) umana, ove il paziente contesti l'ascrivibilità dell'evento di danno al malfunzionamento dell'intelligenza artificiale, la dimostrazione dell'esistenza del nesso causale […] tra "condotta" ed evento dannoso sarà ardua sino, e forse oltre, il limite della *probatio diabolica*».

[82] N. MUCCIOLI, *op. cit*., p. 614; E. COLLETTI, *op. cit.*, p. 269. Sul ricorso al meccanismo presuntivo nel giudizio di responsabilità civile per danni cagionati dall'IA, v., *amplius*, S. CORSO, *op. cit.*, p. 640 ss.; D. MESSNER-KREUZBAUER-J. PEHM, *Taming AI Through Presumptions: a Softer Approach to Tort Law Harmonisation?*, in *ZEuP*, 2024, p. 161 ss.

AI LAW

INTERNATIONAL REVIEW
OF ARTIFICIAL INTELLIGENCE LAW

Ethical Codes in the European Artificial Intelligence Regulation

Valerio Pescatore

Professore ordinario di Diritto privato
nell'Università di Roma "La Sapienza"

(EN): This article examines the role of ethical codes in the new Regulation (EU) 2024/1689 on Artificial Intelligence, situating them within the broader evolution of the relationship between public and private law. Starting from the distinction between self-regulatory codes and codes of conduct, it reconstructs the progressive "juridification" of these instruments and the emergence of regulatory models based on legislative impetus and co-regulation. The analysis distinguishes voluntary codes of conduct, designed to promote ethical principles for AI systems that are not high-risk, from codes of practice, which operate as co-regulatory instruments whose content supplements specific obligations imposed on providers of general-purpose AI models. Particular attention is paid to the European Commission, which encourages the drafting of these codes, assesses their adequacy and, where private initiative is absent or inadequate, may adopt common rules. The article thus highlights a new feature of contemporary law: the interchangeability of public and private roles and the circulation of rules and ordering principles between law and ethics.

(ES): El artículo examina el papel de los códigos éticos en el nuevo Reglamento (UE) 2024/1689 sobre Inteligencia Artificial, situándolos en el contexto más amplio de la evolución de las relaciones entre el Derecho público y el Derecho privado. Partiendo de la distinción entre códigos de autorregulación y códigos de conducta, reconstruye el progresivo proceso de "juridificación" de estos instrumentos y la afirmación de modelos normativos basados en el impulso legislativo y la corregulación. El análisis distingue entre los códigos de conducta voluntarios, destinados a promover principios éticos para sistemas de IA que no sean de alto riesgo, y los códigos de buenas prácticas, instrumentos de corregulación cuyo contenido integra obligaciones específicas impuestas a los proveedores de modelos de IA de uso general. Se presta especial atención a la Comisión Europea, que fomenta la elaboración de estos códigos, evalúa su adecuación y, en caso de inercia o insuficiencia de la iniciativa privada, puede adoptar normas comunes. El trabajo pone así de relieve una nueva fisonomía del Derecho contemporáneo, caracterizada por la intercambiabilidad de los papeles público y privado y por la circulación de reglas y principios ordenadores entre Derecho y ética.

(FR): L'article examine le rôle des codes éthiques dans le nouveau règlement (UE) 2024/1689 sur l'intelligence artificielle, en les situant dans le contexte plus large de l'évolution des rapports entre droit public et droit privé. À partir de la distinction entre codes d'autorégulation et codes de conduite, il retrace le processus progressif de « juridicisation » de ces instruments et l'affirmation de modèles normatifs fondés sur l'impulsion législative et la corégulation. L'analyse distingue les

codes de conduite volontaires, destinés à promouvoir des principes éthiques pour les systèmes d'IA qui ne sont pas à haut risque, des codes de bonnes pratiques, instruments de corégulation dont le contenu complète les obligations spécifiques imposées aux fournisseurs de modèles d'IA à usage général. Une attention particulière est accordée à la Commission européenne, qui encourage l'élaboration de ces codes, en évalue l'adéquation et peut, en cas d'inaction ou d'insuffisance de l'initiative privée, adopter des règles communes. L'étude met ainsi en lumière une nouvelle physionomie du droit contemporain, caractérisée par l'interchangeabilité des rôles public et privé et par la circulation des règles et des principes directeurs entre droit et éthique.

(DE): Der Beitrag untersucht die Rolle ethischer Kodizes in der neuen Verordnung (EU) 2024/1689 über künstliche Intelligenz und ordnet sie in den weiteren Zusammenhang der sich wandelnden Beziehungen zwischen öffentlichem Recht und Privatrecht ein. Ausgehend von der Unterscheidung zwischen Selbstregulierungskodizes und Verhaltenskodizes zeichnet er den fortschreitenden Prozess der „Verrechtlichung" dieser Instrumente sowie die Herausbildung von Regelungsmodellen nach, die auf gesetzgeberischem Impuls und Koregulierung beruhen. Unterschieden werden freiwillige Verhaltenskodizes, die ethische Grundsätze für KI-Systeme fördern sollen, die nicht als Hochrisikosysteme gelten, und Verhaltenskodizes für bewährte Verfahren, die als Instrumente der Koregulierung die besonderen Pflichten der Anbieter von KI-Modellen mit allgemeinem Verwendungszweck ergänzen. Besondere Aufmerksamkeit gilt der Europäischen Kommission, die die Ausarbeitung dieser Kodizes fördert, ihre Angemessenheit bewertet und bei Untätigkeit oder Unzulänglichkeit privater Initiative gemeinsame Vorschriften erlassen kann. Der Beitrag verdeutlicht damit eine neue Gestalt des modernen Rechts, die durch die Austauschbarkeit öffentlicher und privater Rollen sowie durch die Zirkulation von Regeln und Ordnungsprinzipien zwischen Recht und Ethik geprägt ist.

(PT): O artigo examina o papel dos códigos éticos no novo Regulamento (UE) 2024/1689 sobre Inteligência Artificial, situando-os no contexto mais amplo da evolução das relações entre o direito público e o direito privado. Partindo da distinção entre códigos de autorregulação e códigos de conduta, reconstrói o progressivo processo de "juridificação" desses instrumentos e a afirmação de modelos normativos fundados no impulso legislativo e na corregulação. A análise distingue os códigos de conduta voluntários, destinados a promover princípios éticos para sistemas de IA que não sejam de risco elevado, dos códigos de boas práticas, instrumentos de corregulação cujo conteúdo complementa obrigações específicas impostas aos fornecedores de modelos de IA de finalidade geral. Dedica-se especial atenção à Comissão Europeia, que incentiva a elaboração desses códigos, avalia a sua adequação e, em caso de inércia ou insuficiência da iniciativa privada, pode adotar regras comuns. O estudo evidencia, assim, uma nova configuração do direito contemporâneo, caracterizada pela intercambialidade dos papéis público e privado e pela circulação de regras e princípios ordenadores entre direito e ética.

(IT): L'articolo esamina il ruolo dei codici etici nel nuovo Regolamento (UE) 2024/1689 sull'Intelligenza Artificiale, collocandoli nel più ampio contesto dell'evoluzione dei rapporti tra diritto pubblico e diritto privato. Muovendo dalla distinzione tra codici di autodisciplina e codici di condotta, si ricostruisce il progressivo processo di "giuridificazione" di questi strumenti, fino all'affermarsi di modelli normativi fondati sull'impulso legislativo e sulla co-regolazione. Analizzando la disciplina del Regolamento, l'articolo distingue tra codici di condotta volontari, destinati a promuovere principi etici per sistemi di IA non ad alto rischio, e codici di buone pratiche, strumenti di co-regolazione il cui contenuto è diretto a integrare obblighi specifici per i fornitori di modelli di IA per finalità generali. Particolare attenzione è dedicata al ruolo della Commissione europea, che incentiva l'elaborazione di tali codici, ne valuta l'adeguatezza e, in caso di inerzia o inadeguatezza, può adottare norme comuni. Il contributo evidenzia così la nuova fisionomia del diritto contemporaneo,

caratterizzata dall'interscambiabilità dei ruoli tra pubblico e privato e dal commercio di regole e principi tra diritto ed etica.

Summary: 1. Introduction. – 2. From Self-Regulatory Codes to Codes of Conduct. – 3. A Rule-Making Model Based on Legislatively Prompted Codes of Conduct. – 4. Ethical Codes under Regulation (EU) 2024/1689: Codes of Conduct. – 5. Continued: Codes of Practice.

1. Introduction

The subject of so-called ethical codes, viewed through the distinctive lens of the Artificial Intelligence Regulation, once again proves paradigmatic of the current relationship between private and public law.

Regulation (EU) 2024/1689 contains scattered references to such codes: not so much provisions as fragments of provisions. The interpreter's task of deriving systematic guidance from them is therefore far from easy, especially because it does not start from unambiguous concepts.

In seeking to offer a contribution, it is appropriate to begin by comparing the Regulation's at times unsystematic framework with the conclusions that legal scholarship has gradually reached in reflecting on ethical codes – or at least with those conclusions that may be regarded as reasonably settled. The aim is to identify lines of continuity and discontinuity, useful both for understanding the rules on artificial intelligence and for providing an updated classification, precisely in light of the new Regulation, of what we regard as an «ethical code».

2. From Self-Regulatory Codes to Codes of Conduct

As is well known, the expression «ethical code» denotes not a single type but a plural category of instruments: a *genus* whose *species* are not easily isolated and distinguished, because terminological multiplicity is matched not by univocal characteristics but by overlaps and uncertain boundaries [1].

[1] See G. ALPA-T. GALLETTO, *Codici di condotta*, in *Dig. disc. priv.*, *Sez. civ.*, *Agg.*, VI, Torino, 2011, p. 163 ff.; F. BENATTI, *L'evoluzione dei codici etici: dalla funzione promozionale alla valenza giuridica*, in *Banca borsa tit. cred.*, 2015, I, p. 242; L. BALESTRA, *Il ruolo dei codici etici nell'autoregolamentazione del mercato*, in *Giur. comm.*, 2018, I, pp. 216-227; A. SCOTTI,

Thus, whereas the interpreter once encountered the professional codes of ethics, today they must also address codes of self-discipline and self-regulation, codes of behaviour, and codes of conduct. Attempts to define the nuances among these different instruments have not always proved fruitful, as shown by the fact that some of these expressions are used or understood interchangeably [2].

The prevailing account accepts what Natalino Irti described, more than twenty years ago, as the progressive «processo di giuridificazione» [3] of ethical codes.

The problems on which scholars have focused may be reduced to three: the question of content, at the boundary between ethics and law, and consequently the identification of the nature of these instruments; the question of their legal relevance, and therefore the place of ethical codes within the system of sources of law; and the question of their effects – whether and to what extent they are binding, and upon whom: only those who drew them up, those who adhere to them, or also third parties, and which third parties [4].

At the cost of inevitably flattening some nuances, and looking at the present, one may say that, in a first phase, professional codes were regarded, despite the diversity of models, as a manifestation of private self-discipline [5] by private parties or groups of private parties: more precisely, as acts of (collective) private autonomy concerning economic interests, giving rise to obligations, whose binding character is usually conferred by legislation [6].

A second phase may conventionally be said to begin with the introduction of the Consumer Code into the Italian legal system. Some assistance is offered to the domestic interpreter by the definition in Article 18(1)(f) of the Consumer Code, which defines a «codice di condotta» as an agreement or set of rules not imposed by the laws, regulations or administrative provisions of a Member

I codici di condotta tra mercato, impresa e contratto, Milano, 2019, p. 4 ff.; R. SENIGAGLIA, *La vincolatività dei codici etici: ossimoro o sineddoche?*, in *Riv. crit. dir. priv.*, 2011, p. 570, speaks of a «categoria concettuale».

[2] The distinction proposed by N. IRTI, *Due temi di governo societario (responsabilità «amministrativa» - codici di autodisciplina)*, in *Giur. comm.*, 2003, I, p. 693 ff., has remained settled and has repeatedly been taken up by later scholars: between a self-regulatory code, which «regola soprattutto il funzionamento interno della società», and ethical codes, which «nutrono l'incauta ambizione di definire i rapporti dell'impresa con beni e soggetti del mondo esterno, di enunciare fini e obiettivi ultimi», in short, of ennobling the value of profit with «valori» (p. 699).

[3] N. IRTI, *op. loc. ult. cit.*

[4] See L. BALESTRA, *op. cit.*, p. 223 ff.

[5] C. ANGELICI, *Responsabilità sociale dell'impresa, codici etici e autodisciplina*, in *Giur. comm.*, 2011, I, p. 159 ff.; L. BALESTRA, *op. cit.*, p. 222.

[6] See R. SENINGAGLIA, *op. cit.*, p. 567 ff.

State and which defines the behaviour of traders who undertake to be bound by it in relation to one or more particular commercial practices or business sectors.

Two characteristics therefore emerge: a negative one – the code is not imposed by law – and a positive one – its subject matter is the conduct of a trader or undertaking in a particular sector or market.

It may likewise be said, again conventionally, that the third millennium opened a phase in which codes of conduct received ever greater attention from the European legislature. On the one hand, it «encourages and promotes the adoption of codes of conduct as instruments for regulating (...) the relationships established» in a particular market or market segment; on the other, it «incide sulla loro elaborazione e predispone opportune misure di controllo» [7].

3. A Rule-Making Model Based on Legislatively Prompted Codes of Conduct

Thus, professional codes, as expressions of private self-discipline, gave way to codes of conduct intended to induce undertakings to comply with standards of proper behaviour towards the third parties with whom they enter into relations [8]. Alongside – or perhaps in opposition to – codes originating exclusively in private law, to which the legal system accords relevance, stand the codes "encouraged" by the European legislature; in other words, codes «prompted exclusively by voluntary initiative» are joined by codes «ad impulso legislativo» [9].

A few examples.

Article 40 of Regulation (EU) 2016/679 (GDPR) provides that «The Member States, the supervisory authorities, the Board and the Commission shall encourage the drawing up of codes of conduct intended to contribute to the proper application of this Regulation, taking account of the specific features of the various processing sectors and the specific needs of micro, small and medium-sized

[7] A. BELLELLI, *Soft law e co-regolazione*, in R. BOCCHINI (ed.), *Manuale di diritto dell'informatica*, Napoli, 2023, p. 52.

[8] A. SCOTTI, *op. cit.*, p. 6.

[9] R. SENIGAGLIA, *op. cit.*, p. 574. On the distinction between codes originating in private law and codes that «sono attuazione di una delega di produzione normativa, compresi i codici elaborati in forme più o meno intense di coregolamentazione con pubblici poteri e, quindi, i codici di fonte normativa» (...) «emanati dall'autorità amministrativa, in esecuzione, pertanto, di un potere attribuito dalla legge», see also A. SCOTTI, *op. cit.*, p. 7, where it is further stated that the State *imprimatur* provides the basis legitimating such forms of self-regulation to have effects *vis-à-vis* third parties (note 18).

enterprises». Such codes «may» be drawn up by associations and other bodies representing categories of controllers or processors; controllers and processors «not subject to» the Regulation may also «adhere» to them, giving binding commitments to apply them; and they must provide for «mechanisms which enable [...] the mandatory monitoring of compliance with its provisions» by controllers and processors «which undertake to apply it». Following a specific procedure, the supervisory authority «shall approve» the code if it considers that it «provides sufficient appropriate safeguards»; and, after an opinion from the European Data Protection Board, the Commission «may, by way of implementing acts, decide that the approved code [...] ha[s] general validity within the Union».

«General validity within the Union» is a formula which, although not entirely intelligible through traditional doctrinal categories, effectively conveys the nature of these codes of conduct. Initially acts of collective private autonomy, they acquire, as the result of a procedure conducted almost entirely by public bodies, general effect: a «force of law», one might say, for all European citizens.

An act of collective private autonomy becomes a general rule, through a system appropriately described, as is well known, as self-regulation or co-regulation.

Digital transactions are probably the field in which this rule-making model is now most widespread.

Under Regulation (EU) 2019/1150 on online intermediation services, the three paragraphs of Article 17 provide that the Commission «encourage[s]» – the term recurs constantly – «providers of online intermediation services», «providers of online search engines», and «the organisations and associations representing them» to draw up, «together with business users, including SMEs, and the organisations representing them», «codes of conduct (...) intended to contribute to the proper application of this Regulation» or of a specific provision, as well as sector-specific codes of conduct.

Codes of conduct thus become «flexible instruments of social regulation», tailored to the concrete needs of a particular sector; their principal function is «integrare le regole giuridiche poste a livello europeo» [10].

Regulation (EU) 2022/2065 (Digital Services Act) follows the same path: Articles 45 to 47 «encourage and facilitate» the drawing up of voluntary codes of

[10] A. BELLELLI, *op. cit.*, p. 55. R. SENIGAGLIA, *op. cit.*, p. 575, distinguishes among legislatively prompted codes, implementing codes, codes recognised by legislation, codes encouraged by legislation and codes supplementing legislation.

conduct at Union level to contribute to the proper application of the Regulation [11]. The Regulation assigns the Commission and the Board functions of guidance and supervision: prior guidance as to the content of individual codes, which must clearly define their objectives and purposes and take account of the needs and interests of all interested parties; and supervision, including *ex post* supervision, because achievement of the codes' objectives is periodically assessed.

A single provision of that Regulation touches on one of the most delicate issues in the law of codes of conduct. The final paragraph of Article 45 provides that, «[i]n the case of systematic failure to comply with the codes of conduct, the Commission and the Board may invite the signatories to the codes of conduct to take the necessary action».

The problem is plainly the consequences flowing from breach of the provisions of a code of conduct: beyond reputational effects, the issue is what remedies are available to ensure the enforceability of those provisions. In this respect, a mere «invitation» by EU public bodies does not appear particularly effective or capable of ensuring effective protection for injured parties.

4. Ethical Codes under Regulation (EU) 2024/1689: Codes of Conduct

It is against this general background and on these theoretical premises that the provisions on ethical codes in Regulation (EU) 2024/1689 on Artificial Intelligence must be understood.

The relevant, sporadic provisions are found in several recitals and a few articles. They nevertheless regulate two different instruments: codes of practice (Article 56) and codes of conduct (Article 95).

In the absence of definitions, the use of two different expressions permits the interpreter to ask whether there is an ontological distinction between a code of conduct and a code of practice. Were it not for the fact that Article 66(e)(i) mentions both among the tasks of the Artificial Intelligence Board – which may issue recommendations and written opinions, inter alia, «on the development and application of codes of conduct and codes of practice» – one might, at first sight, suppose that the legislature uses the terms indiscriminately, as synonyms. But the jurist's traditional aspiration to classification requires a different reconstructive path, in search of the differences between the two instruments: two legal names must correspond to two different sets of rules.

[11] A. BELLELLI, *op. cit.*, p. 56.

The Regulation deals with codes of conduct in Recitals 20, 27, 165 and 174, and in the corresponding Article 95.

Under paragraph 1, Article 95 encourages and facilitates the drawing up of codes of conduct intended to promote the voluntary application to AI systems other than high-risk AI systems of certain more stringent requirements laid down for high-risk systems, which are not mandatory for the former.

Under paragraph 2, Article 95 facilitates the drawing up of codes of conduct concerning the voluntary application of certain specific requirements to all AI systems without distinction, on the basis of clear objectives and «key performance indicators» capable of measuring their achievement.

Voluntary codes of conduct «may» be drawn up by individual providers or deployers of AI systems, or by organisations representing them, with the participation of any interested stakeholders (paragraph 3); they must take particular account of the interests and specific needs of SMEs and start-ups (paragraph 4).

Recitals 20 and 27 clarify the rationale of these provisions. Codes of conduct are linked to the need for market operators to possess essential AI literacy (Recital 20); the dissemination of such knowledge is likewise expressly connected with «seven ethical principles» which, although described as «non-binding», are treated as a central point of reference «to help ensure that AI is trustworthy and ethically sound», «consistent [...] and human-centric» (Recital 27). Here we are at the heart of the tendency towards the «eticizzazione del fare» [12], which, together with the progressive juridification of codes described above, has blurred the boundaries between ethics and law. This is made still clearer by the list of the seven principles, among which the Regulation mentions «privacy and data governance», «diversity», «non-discrimination and fairness», and «societal and environmental well-being».

Ultimately, the codes of conduct regulated by the AI Act continue the «tradition» of ethical codes: voluntary acts of collective private autonomy, encouraged by the European Union, through which market operators are indirectly called upon to engage in conduct that is not legally required but is ethically desirable.

5. Continued: Codes of Practice

By contrast, Recitals 116, 117, 135 and 179, together with Articles 53 and, above all, 56, are devoted to codes of practice.

[12] N. IRTI, *op. loc. ult. cit.*

Beyond the *nomen*, the Regulation's provisions appear to establish a legal instrument which, while following the line of EU legislative interventions since 2005 mentioned above, presents some partly new features.

Article 56, using the now well-established formula «encourage and facilitate», calls for the «drawing-up» of codes of practice «at Union level», and thus with general scope, «in order to contribute to the proper application of this Regulation» (paragraph 1): so far, even in its literal wording, there is nothing new.

Under paragraphs 2 and 4, the AI Office and the Board «shall aim to ensure that the codes of practice cover at least» certain «obligations», namely those set out in Articles 53 and 55 (paragraph 2). The codes should therefore indicate, with regard to the obligations of providers of general-purpose AI models (Article 53), how to draw up and keep up to date the model's technical documentation, share the information required for its integration into other AI systems, publish a sufficiently detailed summary of the content used for training, and comply with copyright law. In addition, as regards the obligations of providers of general-purpose AI models with systemic risk (Article 55), codes of practice should also provide guidance on how to conduct model evaluations, mitigate risks, report incidents and adopt corrective measures, as well as ensure cybersecurity [13].

Codes of practice should also «clearly set out their specific objectives and contain commitments and measures» (Article 56(4)): here the idea of voluntariness that, as we have seen, informs the provisions on codes of conduct appears to fade, or at least to be diluted.

Under paragraph 6, the Commission «may, by means of an implementing act, approve a code of practice and give it a general validity within the Union». The formula and regulatory mechanism of the 2016 GDPR reappear.

The mechanism, however, evolves in paragraph 9. That paragraph first sets a precise and imminent deadline: «The codes of practice shall be ready by 2 May 2025 at the latest». It then provides that, if by 2 August 2025 a code of practice has not been finalised, or if, following assessment by the AI Office, it is not considered «adequate» because it does not «cover» at least the obligations under Articles 53 and 55, the Commission «may provide, by means of implementing acts, common rules for the implementation» of those obligations.

Of immediate relevance is the publication, on 14 November 2024, of the «Primo progetto del codice di buone pratiche sull'IA per finalità generali».

Still more significant is the legal and systemic point.

[13] See E. MAESTRI (ed.), *Introduzione e note di sintesi al Regolamento (UE) 2024/1689 sull'intelligenza artificiale*, Napoli, 2024, pp. 266, 269, 270 ff.

Codes of practice are conceived as an «essential tool» (Recital 117) through which providers of general-purpose AI models will be led, directly or indirectly, to comply with the obligations laid down by the Regulation.

A felicitous expression used in the preamble to the «Primo progetto» explains that they are to be understood as «a bridge» until harmonised standards for general-purpose AI models are adopted.

In this sense, the AI Act may be regarded as a new «step» on the ladder of self-regulation or, if preferred, as a new and more intensive form of co-regulation. It also performs a temporary function, under a framework that does not overlook either private parties' inaction or an unsuccessful exercise of their autonomy.

The system established by the Regulation provides that codes of practice, encouraged by the Commission – which is, moreover, proceeding in an almost dirigiste manner – are developed with the broad contribution of market operators, namely providers of AI systems. It also provides that, where their content is considered adequate, the codes are approved by the Commission, which gives them «general validity», that is, general effect throughout the Union. This plainly gives them a normative and public character.

The distinctive feature, however, is the alternative: if this form of co-regulation fails to produce a concrete result because the codes are not completed in time or are not approved by the Commission as «adequate», the Commission may then «establish common rules for the implementation of the relevant obligations» (Recital 117 and Article 56(9)).

The «cycle» of rule-making for codes of practice can therefore be summarised as follows: «encouragement» by public authority; exercise of collective private autonomy; assessment of the resulting product by public authority; where the assessment is positive, the autonomous instrument acquires «general validity»; where it is negative, or private autonomy is not exercised, public authority lays down «common rules».

The adoption of codes of practice is, moreover, destined to have a significant impact on the day-to-day operations of providers of general-purpose AI models.

The identical rules laid down in Articles 53(4) and 55(2) provide that, until a «harmonised» technical standard is published [14], providers – including providers of models with systemic risk – may demonstrate compliance with the obligations under those provisions of the Regulation by «relying» on codes of

[14] The «harmonised standard» within the meaning of Article 2(1)(c) of Regulation (EU) 2012/1025 on European standardisation, defined as a «technical specification» adopted, at the Commission's request, by a recognised standardisation body, is expressly referred to in the definition in Article 3(27) of Regulation (EU) 2024/1689.

practice: that is, by proving that they have aligned, or at least modelled, their actual conduct on the codes. The content of the codes thereby becomes an objective benchmark for assessment.

By contrast, consistency of that conduct with the requirements of any harmonised standard that may be adopted gives rise to an actual presumption of conformity with the Regulation's obligations. Providers that have not adhered to an approved code of practice, or have not aligned their conduct with any approved harmonised standard, can demonstrate compliance with the obligations imposed by Articles 53 and 55 only through «alternative adequate means».

If this is so, this complex framework – which ultimately creates a kind of hierarchy among the instruments available to providers for demonstrating compliance with the obligations imposed by the Regulation – suggests two concluding observations.

The first is the strong incentive for private parties to exercise their autonomy, also in accordance with the principle of horizontal subsidiarity. From this perspective, the contribution made by private parties to self-regulation through codes of practice may readily be classified among sources of secondary law [15].

The second observation concerns the supplementary function of public authority, entrusted to the Commission where collective private autonomy is not exercised in time, or is exercised in a manner that the competent bodies consider inadequate. In that event, the European legislature once again confers law-making power on a public body: the Commission.

In this respect too, codes of practice are more an expression of co-regulation than of self-regulation, because the final determination of the obligations in question entails a decisive step within the sphere of public authority: it either approves the product of collective private autonomy – which it has first encouraged and then directed – or may adopt common rules capable of acquiring general and abstract force.

Private autonomy is thus almost hemmed in by public authority: as regards initiative, the latter gives it the impetus and incentive supplied by legislation; as regards the final outcome, it nevertheless subjects it to scrutiny capable of conferring the «general validity» described above. Thus, even within forms of joint participation in the creation of legal rules, the relationship between public power and private rule assumes partly new characteristics [16].

Overall, then, as noted at the outset, this framework vividly expresses, in a sensitive field such as artificial intelligence, a distinctive feature of contempo-

[15] See A. BELLELLI, *op. cit.*, p. 54.

[16] See A. SCOTTI, *op. cit.*, p. 22.

rary law: the «new paradigm» discerned in the «displuvio che caratterizza» today «l'opposizione tra diritto pubblico e diritto privato», a paradigm inspired, among other things, by the «interscambiabilità dei ruoli» and the «commercio delle regole e dei principi ordinatori» [17].

[17] A. ZOPPINI, *Il diritto privato e i suoi confini*, Bologna, 2020, pp. 82-83, which also refers to S. CASSESE, *L'arena pubblica. Nuovi paradigmi per lo Stato*, in *Riv. trim. dir. pubbl.*, 2001, p. 607.