

AMENDED IN SENATE JUNE 17, 2024
AMENDED IN ASSEMBLY MAY 02, 2024
AMENDED IN ASSEMBLY APRIL 22, 2024

CALIFORNIA LEGISLATURE— 2023–2024 REGULAR SESSION

ASSEMBLY BILL

NO. 2013

Introduced by Assembly Member Irwin

January 31, 2024

An act to add Title 15.2 (commencing with Section 3110) to Part 4 of Division 3 of the Civil Code, relating to artificial intelligence.

LEGISLATIVE COUNSEL'S DIGEST

AB 2013, as amended, Irwin. Artificial intelligence: training data transparency.

Existing law requires the Department of Technology, in coordination with other interagency bodies, to conduct, on or before September 1, 2024, a comprehensive inventory of all high-risk automated decision systems, as defined, that have been proposed for use, development, or procurement by, or are being used, developed, or procured by, state agencies, as defined.

This bill would require, on or before January 1, 2026, and before each time thereafter that an artificial intelligence system or service, as defined, is made available to Californians for use, regardless of whether the terms of that use include compensation, a developer of the system or service to post on the developer's internet website documentation, as specified, regarding the data used to train the artificial intelligence system or ~~service~~. *service, including, among other requirements, a high-level summary of the datasets used in the development of the system or service, as specified. For an artificial intelligence system or service made available for use before January 1, 2025, the bill would instead require the high-level summary to use information reasonably available to the developer and would require a developer who is unable to locate information for a high-level summary to post a description of the methods used to search for information, as specified.*

Digest Key

Vote: majority Appropriation: no Fiscal Committee: no Local Program: no

Bill Text

THE PEOPLE OF THE STATE OF CALIFORNIA DO ENACT AS FOLLOWS:

SECTION 1. Title 15.2 (commencing with Section 3110) is added to Part 4 of Division 3 of the Civil Code, to read:

TITLE 15.2. Artificial Intelligence Training Data Transparency

3110. For purposes of this title, the following definitions shall apply:

(a) “Artificial intelligence” means an engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments.

(b) “Developer” means a person, partnership, state or local government agency, or corporation that designs, codes, or produces an artificial intelligence system or service, or substantially modifies an artificial intelligence system or service for use by a third party for free or for a fee. *For purposes of this subdivision, “third party” does not include an affiliate as defined in subparagraph (A) of paragraph (1) of subdivision (c) of Section 1799.1a, or a hospital’s medical staff member.*

(c) “Synthetic data generation” means a process in which seed data are used to create artificial data that have some of the statistical characteristics of the seed data.

(d) “Train an artificial intelligence system or service” includes testing, validating, or fine tuning *by the developer of* the artificial intelligence system or service.

3111. On or before January 1, 2026, and before each time thereafter that an artificial intelligence system or service is made *publicly* available to Californians for use, regardless of whether the terms of that use include compensation, the developer of the system or service shall post on the developer’s internet website documentation regarding the data used *by the developer* to train the artificial intelligence system or service, including, but not be limited to, all of the following:

(a) A high-level summary of the datasets used in the development of the *artificial intelligence* system or service, including, but not limited to:

(1) The sources or owners of the datasets.

(2) A description of how the datasets further the intended purpose of the *artificial intelligence* system or service.

(3) The number of data points included in the ~~datasets~~ *datasets, which may be in general ranges, and* with estimated figures for dynamic datasets.

(4) A clear definition of each category associated to data points within the datasets, including the format of data points and sample values.

(5) Whether the datasets include any data protected by copyright, trademark, or patent, ~~requiring the purchase or licensure of the data,~~ or whether the datasets are entirely in the public domain.

(6) Whether the datasets were purchased or licensed by the developer.

(7) Whether the datasets include personal information, as defined in subdivision (v) of Section 1798.140.

(8) Whether the datasets include aggregate consumer information, as defined in subdivision (b) Section 1798.140.

(9) A description of any cleaning, processing, or other modification to the datasets by the developer, including the intended purpose of those efforts in relation to the *artificial intelligence* system or service. ~~If datasets have been merged with other datasets, the developer shall include the disclosures required by this section for the original datasets.~~

(10) The time period during which the data in the datasets were collected, including a notice if the data collection is ongoing.

(11) The dates the datasets were first ~~and last~~ used during the development of the *artificial intelligence* system or service.

~~(b) A disclosure of whether~~

(12) Whether the artificial intelligence system or service used or continuously uses synthetic data generation in its development. A developer may include a description of the functional need or desired purpose of the synthetic data in relation to the intended purpose of the system or service.

~~(e)~~

(b) A developer shall not be required to post documentation regarding the data used to train an artificial intelligence system or service ~~that has the sole~~ for any of the following:

(1) An artificial intelligence system or service whose sole purpose is to help ensure security and integrity as defined in subdivision (ac) of Section ~~1798.140~~ 1798.140.

(2) An artificial intelligence system or service whose sole purpose is the operation of aircraft in the national airspace.

(3) An artificial intelligence system or service developed for national security, military, or defense purposes that is made available only to a federal entity.

(c) Notwithstanding subdivision (a), for an artificial intelligence system or service made available to Californians for use before January 1, 2025, the high-level summary posted pursuant to subdivision (a) shall use information reasonably available to the developer. A developer who is unable to locate information for a high-level summary shall post a description of the methods used to search for information regarding the data used to train the artificial intelligence system or service, and shall state that the information is not reasonably available.