AB-412 Generative artificial intelligence: training data: copyrighted materials. (2025-2026)

Date Published: 05/07/2025 09:00 PM

AMENDED IN ASSEMBLY MAY 07, 2025

AMENDED IN ASSEMBLY APRIL 28, 2025

AMENDED IN ASSEMBLY APRIL 21, 2025

AMENDED IN ASSEMBLY MARCH 20, 2025

AMENDED IN ASSEMBLY MARCH 10, 2025

AMENDED IN ASSEMBLY FEBRUARY 25, 2025

CALIFORNIA LEGISLATURE— 2025–2026 REGULAR SESSION

ASSEMBLY BILL   NO. 412

Introduced by Assembly Member Bauer-Kahan
(Coauthor: Assembly Member Kalra)

February 04, 2025

An act to add Title 15.3 (commencing with Section 3115) to Part 4 of Division 3 of the Civil Code, relating to artificial intelligence.

LEGISLATIVE COUNSEL'S DIGEST

AB 412, as amended, Bauer-Kahan. Generative artificial intelligence: training data: copyrighted materials.

Existing federal law, through copyright, provides authors of original works of authorship, as defined, with certain rights and protections. Existing federal law generally gives the owner of the copyright the right to reproduce the work in copies or phonorecords and the right to distribute copies or phonorecords of the work to the public. Existing federal law provides that sound recordings fixed before February 15, 1972, are not subject to copyright, but are subject to similar rights and protections under the Classics Protection and Access Act.

Existing law requires, on or before January 1, 2026, and before each time thereafter that a generative artificial intelligence system or service, as defined, or a substantial modification to a generative artificial intelligence system or service, released on or after January 1, 2022, 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 generative artificial intelligence system or service.

This bill would require a developer of a generative artificial intelligence model to, among other things, document any covered materials that the developer knows were used *by the developer* to train the model. The bill would require the developer to make available a mechanism on the developer's internet website allowing a rights owner to submit a request for information about the developer's use of covered materials that would allow the rights owner to provide the developer with, among other things, registration, preregistration, or index numbers and fingerprints for one or more covered materials. The bill ~~would~~ *would, subject to specified exceptions,* require a developer to, within ~~7~~ *30* days of receiving that request from the rights owner, assess whether the covered material represented by a fingerprint provided by the rights owner is likely to be present in the developer's dataset and provide the rights owner with a list of their covered materials that were used to train the model and are likely to be present in the developer's dataset, as specified. The bill would provide that each day following the ~~7-day~~ *30-day* period that a developer fails to provide a rights owner with that information constitutes a discrete violation. The bill would authorize a rights owner *who complies with specified requirements for submitting a request* that is not provided with information according to these provisions to bring a civil action against the developer for specified relief. The bill would provide that the bill's requirements do not apply to a ~~developer that makes all of the data used to train the~~ model *that meets certain criteria, including, among other things, being trained exclusively using data the developer makes* publicly available at no cost, as specified. The bill would define various terms for these purposes.

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

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

SECTION 1. Title 15.3 (commencing with Section 3115) is added to Part 4 of Division 3 of the Civil Code, to read:

TITLE 15.3. Copyrighted Materials Used for Artificial Intelligence Training

3115. For the purposes of this title, the following definitions apply:

(a) "Approximate content fingerprint" or "fingerprint" means an abstract representation of digital content that encodes distinctive features of the content and that is all of the following:

(1) Distinct to the digital content being represented.

(2) Robust to minor variations in the original digital content.

(3) Incapable of being used to reconstruct the original digital content.

(4) Capable of being used to readily identify digital content in a dataset.

(b) "Artificial intelligence" or "AI" 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.

(c) "Covered material" means a material registered, preregistered, or indexed with the United States Copyright Office pursuant to Title 17 of the United States Code, Public Law 94-553 (17 U.S.C. Sec. 101 et seq.).

(d) "Rights owner" means either of the following:

(1) The owner of a copyright enforceable under the copyright laws of the United States pursuant to Title 17 of the United States Code, Public Law 94-553 (17 U.S.C. Sec. 101 et seq.).

(2) The owner of a sound recording fixed before February 15, 1972, enforceable under Title 17 of the United States Code (17 U.S.C. Sec. 1401).

(e) "Developer" means a *business,* person, partnership, corporation, or other entity that designs, codes, produces, or substantially modifies a GenAI model and that does either of the following:

(1) Uses the GenAI model commercially in California.

(2) Makes the GenAI model available to Californians for use.

(f) "Generative artificial intelligence" or "GenAI" means an artificial intelligence system that can generate derived synthetic content, including text, images, video, and audio, that emulates the structure and characteristics of the system's training data.

3116. A developer of a GenAI model shall do all of the following:

(a) (1) Document any covered materials that the developer knows were used *by the developer* to train the GenAI model.

(2) Make reasonable efforts to identify and document any other covered materials that were used *by the developer* to train the GenAI model.

(3) Document the rights owner of each covered material documented pursuant to this subdivision.

(b) *(1)* Make available information on the developer's internet website sufficient to enable a natural person to generate a fingerprint that is both of the following:

~~(1)~~

(A) Compatible with any covered materials used by the developer to train the GenAI model.

~~(2)~~

(B) Generated according to widely accepted industry standards.

(2) The obligation to make available information pursuant to this subdivision may be satisfied by directing rights owners to an external tool that is free to use, nondiscriminatory, and reasonably accessible.

(c) (1) Make available a mechanism on the developer's internet website allowing a rights owner to submit a request for information about the developer's use of covered materials.

(2) The mechanism made available pursuant to this subdivision shall allow a rights owner to provide the developer with all of the following:

(A) Documentation sufficient to establish the rights owner's identity.

(B) The physical or electronic signature of the rights owner or a third party authorized to act on behalf of the rights owner.

(C) Registration, preregistration, or index numbers and fingerprints for one or more covered materials.

(d) Document any requests received using the mechanism established pursuant to subdivision (c).

(e) Retain ~~any~~ *the* documentation required ~~by this section~~ *under subdivisions (a) and (d)* for as long as the developer uses the GenAI model commercially in California or makes the GenAI ~~system or~~ model available to Californians for use, whichever is longer, plus ~~10~~ *five* years.

3117. (a) Within ~~seven~~ *30* days of receiving a request for information from a rights owner using the mechanism established pursuant to subdivision (c) of Section 3116, a developer shall do both of the following:

(1) (A) For each fingerprint provided by the rights owner, assess whether the covered material represented by the fingerprint is likely to be present in the developer's dataset.

(B) A developer shall not be required to assess a fingerprint that was not generated according to widely accepted industry standards.

(2) Provide the rights owner with the following information:

(A) (i) A list of covered materials held by the rights owner that the developer documented pursuant to subdivision (a) of Section 3116.

(ii) A rights owner shall not be required to provide a registration number, preregistration number, index number, or fingerprint to a developer in order to receive the information required under this subparagraph.

(B) A list of covered materials held by the rights owner that a fingerprint assessment suggests are likely to be present in the developer's dataset pursuant to paragraph (1).

(b) A developer's collection, use, retention, and sharing of information from a rights owner pursuant to this section shall be reasonably necessary and proportionate to achieve the purposes for which the information was collected and processed, or for another disclosed purpose that is compatible with the context in which the information was collected, and not further processed in a manner that is incompatible with those purposes.

(c) Each day after the ~~seven day~~ *30-day* period described in subdivision (a) that a developer fails to provide a rights owner with the information required under this title constitutes a discrete violation.

(d) A developer shall not be required to respond to a request that is either of the following:

(1) Not accompanied by documentation sufficient to establish the rights owner's identity.

(2) Made in violation of Section 3118.

3118. *(a) A rights owner, or any person acting on their behalf, shall not submit more than one request per calendar quarter to the same developer concerning the same GenAI model, unless the subsequent request includes material new information not available to the rights owner at the time of the prior request.*

(b) A request submitted pursuant to this section may pertain to multiple covered materials.

~~**3118.**~~***3119.*** A rights owner that *has complied in good faith with Section 3118 and that* is not provided with the information as required by this title may bring a civil action against the developer for any of the following:

(a) One thousand dollars ($1,000) per violation or actual damages, whichever is greater.

(b) Injunctive or declaratory relief.

(c) Reasonable attorney's costs and fees.

(d) Any other relief the court deems appropriate.

~~**3119.**~~***3119.5.*** This title ~~does~~ *shall* not apply to a ~~developer that makes all of the data used to train the developer's~~ GenAI model ~~publicly~~ *that is any of the following:*

(a) Trained exclusively using data the developer makes publicly available at no cost to users of the developer's internet website.

(b) Developed and used exclusively for noncommercial academic or governmental research.

(c) Not trained using covered materials.

(d) Trained exclusively using covered materials for which the developer is the rights owner.