

# STATE OF NEW YORK

6955--A

2025-2026 Regular Sessions

## IN SENATE

March 27, 2025

Introduced by Sen. GOUNARDES -- read twice and ordered printed, and when printed to be committed to the Committee on Internet and Technology -- recommitted to the Committee on Internet and Technology in accordance with Senate Rule 6, sec. 8 -- committee discharged, bill amended, ordered reprinted as amended and recommitted to said committee

AN ACT to amend the general business law, in relation to establishing the artificial intelligence training data transparency act

The People of the State of New York, represented in Senate and Assembly, do enact as follows:

1 Section 1. The general business law is amended by adding a new article  
2 44-C to read as follows:

### ARTICLE 44-C

#### ARTIFICIAL INTELLIGENCE TRAINING DATA TRANSPARENCY ACT

##### Section 1430. Short title.

##### 1431. Definitions.

##### 1432. Data used to train generative artificial intelligence models or services.

9 § 1430. Short title. This act shall be known and may be cited as the  
10 "artificial intelligence training data transparency act".

11 § 1431. Definitions. For the purposes of this article, the following  
12 terms shall have the following meanings:

13 1. "Artificial intelligence" or "artificial intelligence technology"  
14 means a machine-based system that can, for a given set of human-defined  
15 objectives, make predictions, recommendations, or decisions influencing  
16 real or virtual environments, and that uses machine- and human-based  
17 inputs to perceive real and virtual environments, abstract such percep-  
18 tions into models through analysis in an automated manner, and use model  
19 inference to formulate options for information or action.

20 2. "Developer" means a person, partnership, state or local government  
21 agency, or corporation that designs, codes, produces, or substantially

EXPLANATION--Matter in italics (underscored) is new; matter in brackets  
[-] is old law to be omitted.

LBD07975-07-6

1 modifies an artificial intelligence model or service for use by members  
2 of the public.

3 3. "Generative artificial intelligence" means a class of AI models  
4 that emulate the structure and characteristics of input data to generate  
5 derived synthetic content, including, but not limited to, images, vide-  
6 os, audio, text, and other digital content.

7 4. "Substantially modifies" or "substantial modification" means a new  
8 version, new release, or other update to a generative artificial intel-  
9 ligence model or service that materially changes its functionality or  
10 performance, including the results of retraining or fine tuning.

11 5. "Synthetic data generation" means a process in which seed data is  
12 used to create artificial data that have some of the statistical charac-  
13 teristics of the seed data.

14 6. "Train a generative artificial intelligence model or service"  
15 includes testing, validating, or fine tuning by the developer of the  
16 artificial intelligence model or service.

17 7. "Aggregate consumer information" means information that relates to  
18 a group of consumers, from which individual consumer identities have  
19 been removed, that is not linked or reasonably linkable to any consumer  
20 or household, including via a device. Aggregate consumer information  
21 does not mean one or more individual consumer records that have been  
22 de-identified.

23 8. "AI model" means an information system or component of an informa-  
24 tion system that implements artificial intelligence technology and uses  
25 computational, statistical, or machine-learning techniques to produce  
26 outputs from a given set of inputs.

27 § 1432. Data used to train generative artificial intelligence models  
28 or services. 1. On or before January first, two thousand twenty-seven,  
29 and prior to each time thereafter that a generative artificial intelli-  
30 gence model or service, or a substantial modification to a generative  
31 artificial intelligence model or service, released on or after January  
32 first, two thousand twenty-two, is made publicly available to New York-  
33 ers for use, regardless of whether the terms of such use include compen-  
34 sation, the developer of such model or service shall post on the devel-  
35 oper's website documentation regarding the data used by the developer to  
36 train the generative artificial intelligence model or service, including  
37 a high-level summary of the datasets used in the development of the  
38 generative artificial intelligence model or service, including, but not  
39 limited to:

40 (a) the sources or owners of the datasets;

41 (b) a description of how the datasets further the intended purpose of  
42 the artificial intelligence model or service;

43 (c) the number of data points included in the datasets, which may be  
44 in general ranges, and with estimated figures for dynamic datasets;

45 (d) a description of the types of data points within the datasets. For  
46 purposes of this paragraph, the following definitions apply:

47 (i) as applied to datasets that include labels, "types of data points"  
48 means the types of labels used; and

49 (ii) as applied to datasets without labeling, "types of data points"  
50 refers to the general characteristics;

51 (e) whether the datasets include any data protected by copyright,  
52 trademark, or patent, or whether the datasets are entirely in the public  
53 domain;

54 (f) whether the datasets were purchased or licensed by the developer;

1 (g) whether the datasets include personal information or personal  
2 identifying information, as defined in section eight hundred ninety-  
3 nine-aaa of this chapter;

4 (h) whether the datasets include aggregate consumer information;

5 (i) whether there was any cleaning, processing, or other modification  
6 to the datasets by the developer, including the intended purpose of  
7 those efforts in relation to the artificial intelligence model or  
8 service;

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

11 (k) the dates the datasets were first used during the development of  
12 the artificial intelligence model or service; and

13 (l) whether the generative artificial intelligence model or service  
14 used or continuously uses synthetic data generation in its development.  
15 A developer may include a description of the functional need or desired  
16 purpose of the synthetic data in relation to the intended purpose of the  
17 model or service.

18 2. A developer shall not be required to post documentation regarding  
19 the data used to train a generative artificial intelligence model or  
20 service for any of the following:

21 (a) A generative artificial intelligence model or service whose sole  
22 purpose is the operation of aircraft in the national airspace; or

23 (b) A generative artificial intelligence model or service developed  
24 for national security, military, or defense purposes that is made avail-  
25 able only to a federal entity.

26 § 2. This act shall take effect immediately.