

STATE OF NEW YORK

8331--A

2025-2026 Regular Sessions

IN SENATE

June 3, 2025

Introduced by Sen. GONZALEZ -- read twice and ordered printed, and when printed to be committed to the Committee on Rules -- committee discharged, bill amended, ordered reprinted as amended and recommitted to said committee

AN ACT to amend the general business law, in relation to requiring transparency from generative artificial intelligence developers for journalism providers

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

1 Section 1. This act shall be known and may be cited as the "New York
2 artificial intelligence transparency for journalism act".

3 § 2. Legislative findings. The legislature hereby finds and declares
4 that:

5 (a) A free and diverse press was critical in the founding of our
6 democracy and continues to be the lifeblood for a functional society;

7 (b) New York has a compelling interest in protecting news publishers
8 and broadcasters that report and distribute news from unfair business
9 practices and competition. Every day, journalism plays an essential role
10 in New York and in local communities, and the ability of local news
11 organizations to continue to provide the public with critical informa-
12 tion about their communities and enabling news publishers and broadcast-
13 ers to receive fair market value for their content that is used by
14 others will preserve and ensure the sustainability of local and diverse
15 news outlets;

16 (c) Communities without newspapers and broadcast news programs lose
17 touch with government, business, education, and neighbors. They operate
18 without journalists working to keep them informed, uncover truth, expose
19 corruption, and share common goals and experiences;

20 (d) Quality journalism is key to sustaining civic society, strengthen-
21 ing communal ties, and providing information at a deep level;

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

LBD13206-04-5

1 (e) Seventy-three percent of United States adults surveyed said they
2 have confidence in their local newspaper. Broadcasting remains a domi-
3 nant and trusted source of news in communities throughout New York;

4 (f) Studies show that news content comprises a disproportionate amount
5 of generative artificial intelligence training data. News content is
6 especially valuable to artificial intelligence developers because it is
7 high-quality, professional writing created by human beings;

8 (g) After training, generative artificial intelligence systems contin-
9 ue to access news websites, podcasts, broadcasts and digital platforms
10 in order gain access to fact-checked, accurate and up to date content to
11 produce outputs;

12 (h) The vast majority of generative artificial intelligence developers
13 do not obtain permission or compensate news publishers or broadcast news
14 operations for accessing their websites, podcasts, broadcasts and
15 digital platforms for the purposes of building and operationalizing
16 their AI tools and services, in violation of copyright law, those sites'
17 and platforms' terms of service and express prohibitions and prefer-
18 ences;

19 (i) Maximizing the potential of generative AI requires ensuring the
20 sustainability of journalism and the news industry; and

21 (j) News publishers and broadcast news operations deserve to know when
22 generative artificial intelligence developers have accessed news
23 websites and used their work.

24 § 3. Article 21-A of the general business law is renumbered article
25 21-B and a new article 21-A is added to read as follows:

26 ARTICLE 21-A

27 ARTIFICIAL INTELLIGENCE SOURCE DATA TRANSPARENCY

28 Section 338. Definitions.

29 338-a. Artificial intelligence source data transparency.

30 338-b. Enforcement.

31 338-c. Applicability.

32 338-d. Severability.

33 § 338. Definitions. The following terms, whenever used or referred to
34 in this article, shall have the following meanings:

35 1. "Artificial intelligence" means a machine-based system that can,
36 for a given set of human-defined objectives, make predictions, recommen-
37 dations, or decisions influencing real or virtual environments, and that
38 uses machine and human-based inputs to perceive real and virtual envi-
39 ronments, abstract such perceptions into models through analysis in an
40 automated manner, and use model inference to formulate options for
41 information or action.

42 2. "Access" means to obtain, retrieve, acquire, reproduce, crawl,
43 index, or request and receive a transmission of content.

44 3. "Covered publication" means any print, broadcast, broadcast network
45 or digital publication or service which:

46 a. performs a public-information function comparable to that tradi-
47 tionally served by journalism organizations, such as newspapers, broad-
48 cast news operations, broadcast network news operations, magazines and
49 other periodical publications;

50 b. invests substantial expenditure of labor, skill, and money to
51 create, edit, produce, and distribute content including by engaging
52 natural persons to create, edit, produce, and distribute original text,
53 audio, photo, illustrative, or video content concerning matters or
54 topics of interest or use to members of the public through activities
55 such as observation, video recording events, interviews, research, test-
56 ing, and analysis; and

1 c. publishes new content or updates its content on at least a monthly
2 basis and has a process for error correction and clarification.

3 4. "Crawler" means software that accesses content from a website or
4 other internet source, such as an online crawler, spider, fetcher,
5 client, bot, user agent or equivalent tool.

6 5. "Developer" means a person that designs, codes, produces, or
7 substantially modifies an artificial intelligence system or service for
8 use by members of the public. The term "developer" shall not include a
9 journalism provider that uses, develops or obtains artificial intelli-
10 gence systems solely for internal use.

11 6. "Generative artificial intelligence" means a class of artificial
12 intelligence models that emulate the structure and characteristics of
13 input data to generate derived synthetic content, including, but not
14 limited to, images, videos, audio, text, and other digital content.

15 7. "Journalism provider" means any person that broadcasts or publishes
16 one or more covered publications.

17 8. "Person" means a natural person, corporation, trust, estate, part-
18 nership, incorporated or unincorporated association or any other legal
19 entity.

20 9. "Artificial intelligence utilization" means to use digital content
21 as data to develop the capabilities of a generative artificial intelli-
22 gence system, including through setting or changing its learnable
23 weights and other parameters, and includes, in addition to the initial
24 dataset training, further testing, validating, grounding, or fine tuning
25 by the developer of the artificial intelligence system or service.

26 § 338-a. Artificial intelligence source data transparency. 1. a. On or
27 before January first, two thousand twenty-seven and before each time
28 thereafter that a generative artificial intelligence system or service,
29 or a substantial modification to a generative artificial intelligence
30 system or service released on or after January first, two thousand twen-
31 ty-two, is made publicly available to New Yorkers for use, regardless of
32 whether the system or service is made available for a fee, the developer
33 of the system or service shall post on the developer's internet website
34 the following information regarding video, audio, text and data from a
35 covered publication used to train the generative artificial intelligence
36 system or service:

37 (i) the uniform resource locators or uniform resource identifiers
38 accessed by crawlers deployed by the developer or by third parties on
39 their behalf or from whom they have obtained video, audio, text or data;

40 (ii) a detailed description of the video, audio, text and data from a
41 covered publication used for artificial intelligence utilization,
42 including the type and provenance of the video, audio, text and data and
43 the means by which it was obtained, sufficient to identify individual
44 works;

45 (iii) whether any source identifiers, terms, or copyright notices were
46 removed from the video, audio, text or data; and

47 (iv) the timeframe of data collection.

48 b. The information required to be posted on a developer's internet
49 website pursuant to paragraph a of this subdivision shall not be
50 required where there is an express written agreement authorizing the
51 developer to access the journalism provider's content and the parties
52 agree not to post information relating to the journalism provider's
53 content on the developer's website.

54 2. a. On or before January first, two thousand twenty-seven, the
55 developer of a generative artificial intelligence system or service who
56 deploys a crawler, either directly or through a third party, in

1 connection with such system or service shall disclose information
2 regarding the identity of crawlers used by the developer or by third
3 parties on the developer's behalf in a manner clearly accessible by a
4 website operator, including but not limited to:

5 (i) the name of the crawler including the crawler's IP address, and
6 specific identifier actually used by the crawler when conducting the
7 crawling activity (such as including the identifiers as part of the user
8 agent or other part of the request headers);

9 (ii) the legal entity responsible for the crawler;

10 (iii) the specific purposes for which each crawler is used;

11 (iv) the legal entities to which operators provide data scraped by the
12 crawlers they operate; and

13 (v) a single point of contact to enable third parties whose websites
14 are accessed by such crawlers to communicate with the developer and to
15 lodge complaints.

16 b. The information disclosed pursuant to paragraph a of this subdivi-
17 sion shall be available on an easily accessible platform and updated at
18 the same time as any change is made to such information.

19 c. The exclusion of a crawler by a website operator shall not nega-
20 tively impact the findability of the website operator's content in a
21 search engine.

22 § 338-b. Enforcement. 1. A journalism provider, or a person authorized
23 to act on a journalism provider's behalf, may bring an action in the
24 supreme court for statutory damages or injunctive relief to compel a
25 developer to comply with section three hundred thirty-eight-a of this
26 article. If the court finds that the developer of a generative artifi-
27 cial intelligence system that is made available to New Yorkers for use
28 is not in compliance, the court shall order such compliance and may
29 impose statutory damages to the journalism provider of up to ten thou-
30 sand dollars for each violation.

31 2. Consistent with the New York laws of civil procedure, the supreme
32 court may issue an order for disclosure of copies of, or records suffi-
33 cient to identify with certainty, the text and data used to train the
34 generative artificial intelligence system or service insofar as such
35 text and data pertains to the journalism provider's internet website,
36 broadcasts, podcasts or other digital platforms, including but not
37 limited to:

38 a. the uniform resource locators accessed by crawlers deployed by
39 developers or by third parties on their behalf or from whom they have
40 obtained text, video, audio or data, and dates and times of collection;
41 and

42 b. the text and data used for artificial intelligence utilization,
43 including the type and provenance of the text and data and the means by
44 which such text and data was obtained and when.

45 § 338-c. Applicability. The provisions of this article shall not be
46 construed to modify, impair, expand, or in any way alter rights pertain-
47 ing to Title 17 of the United States Code or the Lanham Act (15 U.S.C.
48 1051 et seq.).

49 § 338-d. Severability. If any provision of this article or the appli-
50 cation thereof to any person or circumstances is held to be invalid,
51 such invalidity shall not affect other provisions or applications of
52 this article which can be given effect without the invalid provision or
53 application, and to this end the provisions of this article are severa-
54 ble.

55 § 4. This act shall take effect immediately.