

STATE OF NEW YORK

8595

2025-2026 Regular Sessions

IN ASSEMBLY

May 22, 2025

Introduced by M. of A. OTIS -- read once and referred to the Committee on Science and Technology

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;

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

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

LBD13206-01-5

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

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

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

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

18 (j) News publishers, broadcast news operations and the public deserve
19 to know when generative artificial intelligence developers have accessed
20 news websites and used their work.

21 § 3. Article 21-A of the general business law is renumbered article
22 21-B and a new article 21-A is added to read as follows:

23 ARTICLE 21-A

24 ARTIFICIAL INTELLIGENCE SOURCE DATA TRANSPARENCY

25 Section 338. Definitions.

26 338-a. Artificial intelligence source data transparency.

27 338-b. Enforcement.

28 338-c. Applicability.

29 338-d. Severability.

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

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

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

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

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

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

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

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

4 5. "Developer" means a person that designs, codes, produces, or
5 substantially modifies an artificial intelligence system or service for
6 use by members of the public. The term "developer" shall not include
7 artificial intelligence systems used, developed or obtained by a jour-
8 nalism provider for internal use.

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

13 7. "Journalism provider" means any person that:

14 a. broadcasts or publishes one or more covered publications; and

15 b. is covered by media liability insurance.

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

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

25 § 338-a. Artificial intelligence source data transparency. 1. a. On or
26 before January first, two thousand twenty-seven and before each time
27 thereafter that a generative artificial intelligence system or service,
28 or a substantial modification to a generative artificial intelligence
29 system or service released on or after January first, two thousand twen-
30 ty-two, is made publicly available to New Yorkers for use, regardless of
31 whether the system or service is made available for a fee, the developer
32 of the system or service shall post on the developer's internet website
33 the following information regarding video, audio, text and data from a
34 covered publication used to train the generative artificial intelligence
35 system or service:

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

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

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

46 (iv) the timeframe of data collection.

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

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

1 regarding the identity of crawlers used by the developer or by third
2 parties on the developer's behalf in a manner clearly accessible by a
3 website operator, including but not limited to:

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

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

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

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

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

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

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

21 § 338-b. Enforcement. 1. a. A journalism provider, or a person author-
22 ized to act on a journalism provider's behalf, may request the clerk of
23 the supreme court, or a judge where there is no clerk, to issue a
24 subpoena to a developer of a generative artificial intelligence system
25 that is made available to New Yorkers for use, regardless of whether the
26 system or service is made available for a fee, for disclosure of copies
27 of, or records sufficient to identify with certainty, the text and data
28 used to train the generative artificial intelligence system or service
29 insofar as such text and data pertains to the journalism provider's
30 internet website, broadcasts, podcasts or other digital platforms,
31 including but not limited to:

32 (i) the uniform resource locators accessed by crawlers deployed by
33 developers or by third parties on their behalf or from whom they have
34 obtained text, video, audio or data, and dates and times of collection;
35 and

36 (ii) the text and data used for artificial intelligence utilization,
37 including the type and provenance of the text and data and the means by
38 which such text and data was obtained and when.

39 b. A subpoena issued pursuant to paragraph a of this subdivision may
40 require disclosure of the information required pursuant to paragraph a
41 of this subdivision in the native form in which such information was
42 copied and stored (including all accompanying keys, values, tags, and
43 the like, and any other available metadata), subject to entry of a suit-
44 able protective order in the case that such information constitutes a
45 trade secret of the generative artificial intelligence system developer.

46 c. The developer shall provide the subpoenaed information within thir-
47 ty days of service of the subpoena or, in the case of trade secrets,
48 entry of a suitable protective order. Such subpoena shall be subject to
49 the provisions of article twenty-three of the civil practice law and
50 rules. The court may impose a penalty for failure to respond to such
51 information subpoenas pursuant to section twenty-three hundred eight of
52 the civil practice law and rules.

53 2. a. A journalism provider may bring an action in the supreme court
54 for an injunction to compel a developer to comply with section three
55 hundred thirty-eight-a of this article.

1 b. If a developer fails to comply with a subpoena issued pursuant to
2 subdivision one of this section, the journalism provider requesting such
3 subpoena may move in the supreme court to compel compliance. If the
4 court finds that the developer did not comply with the subpoena, the
5 court shall order compliance and may impose statutory damages to the
6 journalism provider requesting such subpoena in the sum of not less than
7 ten thousand dollars nor more than fifty thousand dollars for each day
8 on which such noncompliance occurs or continues.

9 c. If the developer fails to comply with a court order issued pursuant
10 to paragraph b of this subdivision, then the journalism provider may
11 request that the attorney general bring an action on their behalf to
12 ensure compliance with the court order and any statutory damages
13 assessed.

14 § 338-c. Applicability. The provisions of this article shall not be
15 construed to modify, impair, expand, or in any way alter rights pertain-
16 ing to Title 17 of the United States Code or the Lanham Act (15 U.S.C.
17 1051 et seq.).

18 § 338-d. Severability. If any provision of this article or the appli-
19 cation thereof to any person or circumstances is held to be invalid,
20 such invalidity shall not affect other provisions or applications of
21 this article which can be given effect without the invalid provision or
22 application, and to this end the provisions of this article are severa-
23 ble.

24 § 4. This act shall take effect immediately.