October 31, 2026

BY EMAIL TO THE PRO SE INTAKE UNIT

The Honorable Sidney H. Stein
United States District Court for the Southern District of New York
Daniel Patrick Moynihan United States Courthouse
500 Pearl Street
New York, NY 10007

**Re: *In re OpenAI, Inc. Copyright Infringement Litigation*, No. 25-md-3143 (SHS) (OTW)**
**This Document Relates To: All Cases**
**Letter Motion for Leave to File Brief of Amicus Curiae Out of Time**

Dear Judge Stein:

Under the Court's Order of August 7, 2026 (ECF No. 1663), as modified by its Order of September 24, 2026 (ECF No. 2049), and Federal Rule of Civil Procedure 6(b)(1)(B), I respectfully move for leave to file, out of time, the attached Exhibit A, "[Proposed] Brief of Amicus Curiae Muhammad Deckri Algamar, AIGP, in Support of Neither Party." A proposed order is attached as Exhibit B.

The deadline under ECF No. 2049 was October 30, 2026. I am an individual in Madrid, I am not admitted to practice in this Court, and I appear without counsel.

The proposed brief addresses a single question that survives either ruling on fair use: how an author can learn whether a work was used to train a model. It contains fewer than 3,000 words, within the 6,000-word limit in ECF No. 1663. Its first footnote gives the statement on authorship and funding.

Respectfully submitted,



Muhammad Deckri Algamar, AIGP
Amicus curiae, appearing pro se
Madrid 28029, Spain
deckrialgamar@student.ie.edu

**EXHIBIT A**

Proposed brief of amicus curiae

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OPENAI, INC., COPYRIGHT INFRINGEMENT LITIGATION

No. 25-md-3143 (SHS) (OTW)

This Document Relates To: All Cases

**[PROPOSED] BRIEF OF AMICUS CURIAE MUHAMMAD DECKRI ALGAMAR, AIGP, IN SUPPORT OF NEITHER PARTY**

## INTEREST OF AMICUS CURIAE

Muhammad Deckri Algamar, AIGP, is an LL.M. candidate at IE University in Madrid who researches copyright and AI governance. He files in his personal capacity, and his views represent no institution. He is not admitted to practice in this Court and appears pro se.[1]

He studies how an author can learn whether a work was used to train an AI model. He is building re(compense), a free resource for authors, still in development and shown in Part III. He has no financial interest in the outcome of this litigation and supports neither party.

## SUMMARY OF ARGUMENT

This brief takes no position on whether training an AI model on a book is fair use. It addresses a question that survives either answer: how does an author find out whether a book was used?

Today the author cannot, short of a lawsuit. Prompting a model shows only what the model retained. Developers have no duty to answer. The most reliable list an author can search, in *Bartz v. Anthropic*, came out of a class action. An author should not need a federal lawsuit to learn whether a book went into a training set.

[1]No party or party's counsel wrote this brief in whole or in part, and no one other than amicus paid for it. Amicus used Claude, an AI tool made by Anthropic PBC, to research and draft it. Anthropic is not a party to this litigation. Amicus has checked every statement and citation and takes responsibility for them.

A check that works without litigation is practical, and Part III describes one. Congress or the developers would have to create it. This Court can still help under either outcome: it can name what the record shows, apply the presumption of public access to the names of works and datasets, and build any class relief on a searchable list.

## ARGUMENT

### I. Today an author cannot find out without suing

#### *A. Asking the model*

A prompt can show that a model retained a text. It cannot show that a book was never used. Memorization rises with repetition, so prompts work best on widely repeated text and least on a book crawled once. *See* Nicholas Carlini et al., *Quantifying Memorization Across Neural Language Models* 2, 5, 10 (ICLR 2023), arXiv:2202.07646. In *Kadrey*, even prompts built to make Meta's models regurgitate drew no more than 50 tokens of the plaintiffs' books. *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026 (N.D. Cal. 2025) (Dkt. 598, at 12-13, 27).

Courts abroad treat this evidence with care. The Munich Regional Court accepted simple prompts where use of the lyrics as training data was undisputed. LG München I, 42 O 14139/24 (Nov. 11, 2025) (*GEMA v. OpenAI*) (not final). At the interim stage the Delhi High Court gave no weight to outputs for articles published after the training cutoff and called a prompt asking for exact words adversarial. *ANI Media Pvt. Ltd. v. Open AI Opco LLC*, I.A. 45300/2024 in CS(COMM) 1028/2024, paras. 82-84, 104 (Delhi HC July 24, 2026).

An author who presses harder meets the developer's terms. OpenAI's terms bar automatic or programmatic extraction of data or output. OpenAI, *Terms of Use* (eff. Jan. 1, 2026), https://openai.com/policies/row-terms-of-use/. OpenAI has told this Court that the Times "paid someone to hack OpenAI's products." Mem. of Law in Supp. of Mot. to Dismiss, *N.Y. Times Co.*

*v. Microsoft Corp.*, No. 1:23-cv-11195, Dkt. 52, at 2 (S.D.N.Y. Feb. 26, 2024). Amicus takes no view on that dispute. It shows how exposed this route is.

***B. Asking the developer***

Amicus has found no law in force that gives an author a work-level answer on request. The EU AI Act requires a public summary of training content, and its template excludes work-level detail. Regulation (EU) 2024/1689, art. 53(1)(d); Commission Notice, C(2025) 8311 final, para. 16 (Dec. 5, 2025). Amicus's coded review covered the 107 summaries online on October 23, 2026. None of the 58 from the nine large developers gives the list of top domain names the template requires. California requires a high-level summary of datasets. Cal. Civ. Code § 3111. The TRAIN Act would give an author a clerk-issued subpoena on a sworn good faith belief. S. 2455, 119th Cong. (2025); H.R. 7209, 119th Cong. (2026). Both bills are in committee.

***C. Reading a list***

Early on, developers named their datasets, and the first suits pleaded those names. OpenAI closed that channel in March 2023, when its GPT-4 technical report gave no further details on dataset construction. OpenAI, *GPT-4 Technical Report* 2 (Mar. 2023), arXiv:2303.08774.

The Atlantic's AI Watchdog searches leaked datasets such as Books3 and LibGen, and shows only that a book sits in one. The most reliable list an author can search came from a lawsuit. The *Bartz* settlement lists 482,460 works, searchable by author, title, publisher, ISBN, ASIN or registration number, and built from files Anthropic produced in discovery. Even that list shows a download and not training: Anthropic represented that neither shadow library was in the training corpus of any commercially released model. *Bartz v. Anthropic PBC*, No. 3:24-cv-05417, Dkt. 680, at 4, 11 (N.D. Cal. July 20, 2026).

This litigation has not produced a list either. Under the inspection protocol, training data is open to counsel and experts and closed to the authors themselves. MDL Dkt. 483 ¶¶ 2, 5, 7(a). A request that OpenAI identify which of the News Plaintiffs' works trained each model was denied with a direction to meet and confer. No. 1:23-cv-11195, Dkt. 365, at 3. A later motion to make OpenAI run searches of its training data was denied. MDL Dkt. 1170, at 2. Amicus found no order requiring OpenAI to name the works it trained on, and does not question these discovery rulings. They show that even plaintiffs with counsel in this Court had to search for themselves. An author outside the case has nothing.

**II. Both answers to the fair use question need the same facts**

If the use is not fair, relief depends on identifying the works. The *Bartz* class was certified on the premise that a per-work list could be built from Anthropic's own records. *Bartz*, Dkt. 244, at 15-16, 25 (N.D. Cal. July 17, 2025). In England a representative claim was refused on the evidence then filed, because membership turned on whether each work had been used in training. *Getty Images (US), Inc. v. Stability AI Ltd.* [2025] EWHC 38 (Ch), [77]-[83], [119].

If the use is fair, the author's questions survive. Fair use attaches to a specific use and not to a technology. *See Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 533 (2023). A ruling for a developer on training leaves open how a copy was acquired, *see Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007 (N.D. Cal. 2025) (Dkt. 231, at 18-24, 31) (rejecting the argument that pirated library copies must be treated as training copies), what a model outputs and what a contract allows. An author cannot test any of those without knowing the work was used.

Disclosure also sits comfortably with fair use. *Google Books*, the leading mass-digitization case in this Circuit, involved a public search: a reader could enter a term and receive "a list of all books in the database in which those terms appear." *Authors Guild v. Google, Inc.*,

804 F.3d 202, 208-09 (2d Cir. 2015). A developer with a fair use to defend loses little by saying which books it used.

The burden points the same way. Fair use is an affirmative defense, and its proponent needs "favorable evidence about relevant markets." *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 590 (1994). A licensing market counts more readily "when the means for paying for such a use is made easier." *Cf. Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 930-31 (2d Cir. 1994). An author cannot price or license a use the author cannot see. Amicus takes no position on what this record shows on that factor.

## III. A check that works without a lawsuit

### *A. How far an author gets today: re(compense)*

Amicus built re(compense) to see how far an author can get without a lawsuit. It opens on the question in this brief and splits it into five facts: a copy acquired, held in a dataset, trained on, retained in a model, reproduced in an output. Each point on its map is something that reaches one of those facts, whether a check the author can run, a request the author can write, or a court or regulator.

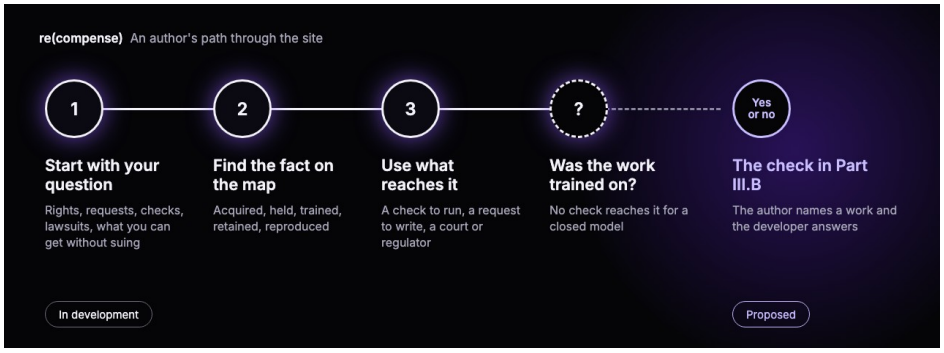



*Figure 1. An author's path through re(compense), and where the proposed check would sit.*

The map behind the second step has one dashed link. Beside "Trained" it reads: "No check reaches it for a closed model." The tools near that link are disclosures, a court order or bills. An author's path through the site ends there.

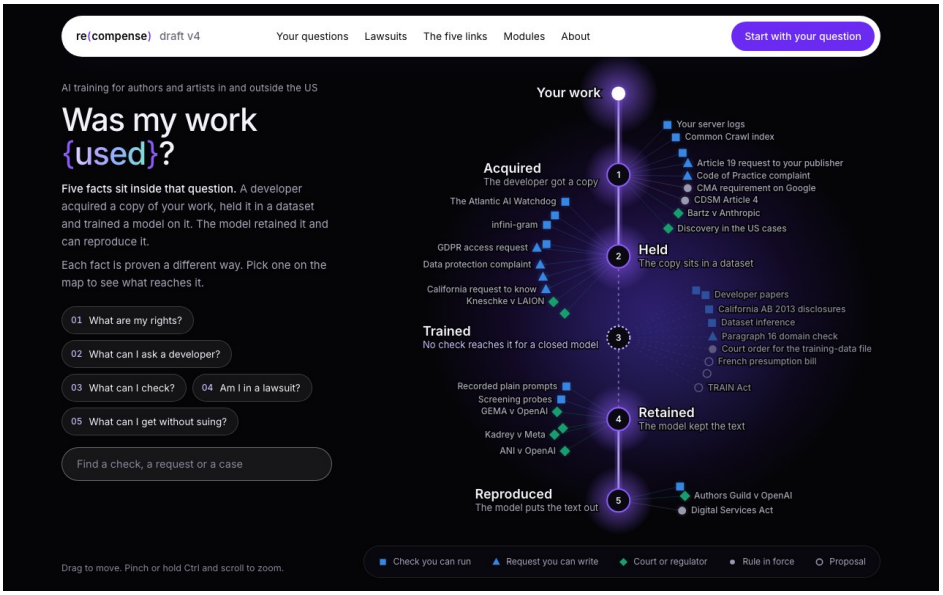



*Figure 2. The re(compense) map, draft v4. The third link is dashed.*

Lawsuits meet the same wall. The module "Am I in a case?" asks three questions about one work and reads the answers against ten pending class definitions. For a work with no US registration whose author lives abroad, nine classes shut the work out and the tenth depends on one more fact. A better answer is still conditional: each class "also needs the developer to have used the work." Only the developer can supply that fact.

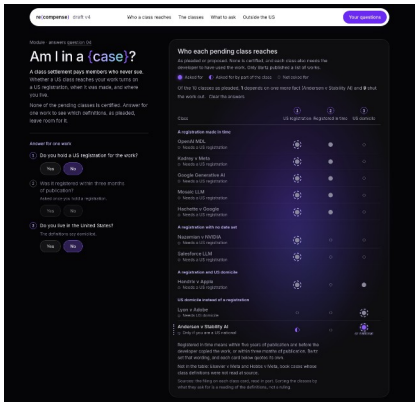



*Figure 3. "Am I in a case?" answered for an unregistered work by an author abroad.*

***B. Three designs that fill the gap***

A published list. A developer publishes the datasets and sources it trained on, down to works where it holds that record.

A yes-or-no check. An author submits a title, ISBN or registration number. The developer checks its records and answers yes, no or cannot tell. Nothing else about the data is disclosed.

A protected plain prompt. An author who types plain prompts about the author's own work, by hand, faces no sanction under the developer's terms.

Any one of them would do. The yes-or-no check is the least intrusive, since it reveals nothing beyond one work the author already knows about. It is also practical. The *Bartz* site runs that search over its list of works, and BigCode's "Am I in The Stack?" runs one for source code by GitHub username. The News Plaintiffs report that OpenAI confirmed it should be feasible to identify which crawls correspond to which training sets. MDL Dkt. 561, at 2, *denied without prejudice*, MDL Dkt. 612. That statement concerns crawls and not works.

Each answer should say which of the five facts it proves. The *Bartz* list answers the first. An answer that names its fact can be read correctly by authors, developers and courts.

This Court cannot create that right. Congress can, and developers can offer it without waiting.

**IV. What this Court can do under either outcome**

First, name what the record shows. Summary judgment papers are judicial documents with a strong presumption of public access. *Lugosch v. Pyramid Co. of Onondaga*, 435 F.3d 110, 121, 123 (2d Cir. 2006). An opinion that names the datasets and sources on the record gives every author something to check.

Second, apply that presumption to names. Omnibus motions to seal summary judgment and *Daubert* exhibits are due January 13, 2027. ECF No. 2101 ¶ 5. Model weights, code and commercial terms may deserve protection. The title of a published book and the name of a dataset are different in kind. Amicus asks that any request to seal them meet the requirement of specific, on-the-record findings. *Lugosch*, 435 F.3d at 120, 124, 126. Judge Chhabria called Meta's request to seal its briefing "preposterous" and ordered the briefs filed unredacted, and his order quotes a Meta employee on a dataset "we know to be pirated, such as LibGen." *Kadrey*, Dkt. 373, at 1-2 (N.D. Cal. Jan. 8, 2025).

Third, build any class relief on a searchable list. If this litigation reaches a class judgment or settlement, the list should be searchable by author, title, ISBN and registration number, and should state which fact it proves.

## CONCLUSION

Whichever way the Court rules on fair use, authors will still ask whether their work was used. The Court should take the three steps in Part IV.

Respectfully submitted,

Dated: October 31, 2026



Muhammad Deckri Algamar, AIGP
Amicus curiae, appearing pro se
Madrid 28029, Spain
deckrialgamar@student.ie.edu

**EXHIBIT B**

Proposed order

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OPENAI, INC., COPYRIGHT INFRINGEMENT LITIGATION

No. 25-md-3143 (SHS) (OTW)

This Document Relates To: All Cases

**[PROPOSED] ORDER**

Upon the letter motion of Muhammad Deckri Algamar, AIGP, for leave to file a brief as amicus curiae out of time, and for good cause shown,

IT IS HEREBY ORDERED that the motion is GRANTED. The Clerk of Court is directed to file the brief attached as Exhibit A to the motion.

Dated: New York, New York
_____, 2026

SO ORDERED:

_____
Sidney H. Stein, U.S.D.J.