DATA, TECH & IP - JUSTICE & LITIGATION

NOVEMBER 11, 2026

Op-Ed: “AI, News, and Copyright: Drawing the Lines after AG Szpunar’s Opinion in Like Company (C-250/25)”

João Pedro Quintais

Advocate General (AG) Szpunar’s Opinion in *Like Company* (C-250/25), delivered on 3 November 2026, offers the Court of Justice its first opportunity to address generative AI and copyright. Its central message is that the legal questions raised by AI cannot be resolved without distinguishing how a model is trained from how a chatbot retrieves and presents information. Although that distinction is persuasive, the Opinion also exposes challenging questions about the protection of press content, the territorial reach of EU law, and the limits of copyright as a response to AI.

The case concerns a claim brought by Hungarian publisher Like Company against Google over the use of its articles in connection with the Gemini chatbot. The national court referred questions about model training and about the reproduction and making available of press content through chatbot responses. In a long Opinion spanning 240 paragraphs, AG Szpunar recommends that the Court address only the latter. Still, he carries out an extensive discussion of AI training, which he offers in the alternative, should the Court of Justice take up these questions (see also the AG’s own explainer).

The Inadmissible Questions: Training, TDM, and Territory

The AG rejects the Commission’s argument that the entire reference is inadmissible. Despite uncertainties about the facts and the technology, he considers that the Court can provide useful guidance on the use of press publications in chatbot responses. In a memorable aside, he

suggests that the referring court is 'evidently lost in the AI landscape' (para. 35).

By contrast, he considers the questions on training to be hypothetical. There is no evidence that Like Company's articles were used to train the relevant Gemini models. In his view, the example before the national court is better explained by the chatbot retrieving an article when answering a query. He also considers that model training is not an 'online use' of press publications within the meaning of Article 15 of Directive 2019/790 (the Copyright in the Digital Single Market Directive, or CDSMD).

This reading matters because it clarifies the narrow scope of the press publishers' right. But it does not place AI training beyond copyright law. In his alternative analysis, AG Szpunar considers that training generally involves reproductions of protected material covered by Article 2 of Directive 2001/29. Those acts must be distinguished from the trained model itself, whose representation of statistical relationships does not ordinarily amount to a reproduction of its training material.

The AG also concludes that training generally falls within the EU law concept of text and data mining (TDM). Article 2(2) CDSMD defines TDM broadly by reference to automated analysis of text and data to generate information, including patterns and trends. In his view, extracting linguistic patterns for training qualifies even if those patterns are subsequently used to generate text. This reading, which is correct in my view (Quintais 2025a), takes a consequential position on a point that has been discussed in scholarship and policy circles for the past years (ECS 2025a).

This does not mean that every training activity is authorised. Article 4 CDSMD requires lawful access and applies only where rights have not been appropriately reserved. AG Szpunar further considers that applying the exception to training generally satisfies the 'three-step test', which limits the scope of copyright exceptions, while leaving open exceptional cases requiring individual assessment (paras 97-152, aligned with e.g. Senftleben 2025). The Opinion thus supports a broad reading of the TDM exception without dispensing with the exception's conditions.

His analysis of memorisation is equally significant. AI models may sometimes reproduce material encountered during training. But, for AG Szpunar, the ability to generate a (partial) copy does not establish that a reproduction of the work is stored inside the model. He considers that inference 'far-fetched' (para. 90). The AG's legal reasoning is persuasive. However, memorisation will often also be a factual question for courts to assess considering the model and circumstances at issue.

In any case, if accepted by the Court, this reasoning could affect national litigation, including the Munich Regional Court's judgments in *GEMA v OpenAI* and *GEMA v Suno*, which the Opinion discusses in this context. It could call into question an important – and flawed– part of the national courts' reasoning on material retained in models. It would not, however, decide whether training copies or specific outputs infringe rights, as those remain distinct questions.

The AG's territorial analysis is also consequential. Making an AI model available in the EU after training abroad does not, in Szpunar's view, automatically bring the foreign training acts within EU copyright law (as argued in e.g. Quintais 2025b). He does not exclude application of EU law to scraping material from servers located in the EU, even if training occurs elsewhere (as suggested by Peukert 2024). Given the uncertainty, he favours a legislative solution respecting territoriality (paras 60-73).

This distinction matters to the ongoing reform debate. EU market access, the location of training, and the location of protected material do not necessarily provide interchangeable connections to EU copyright law. A workable legislative response must identify which acts it regulates and explain the territorial basis for doing so.

The Admissible Questions: Retrieval and the News

The Opinion's most immediate contribution lies in its proposed framework for retrieval and chatbot responses. A chatbot may use *retrieval-augmented generation* (RAG): instead of relying solely on a trained model, it retrieves online material and uses it as context for an answer. AG Szpunar considers that this can constitute an online use of a press publication under Article 15 CDSMD.

He correctly distinguishes RAG from TDM for training (ECS 2026). In the process he examines, the chatbot retrieves specific content to improve an answer rather than analysing data to extract patterns or other information. The technical copies involved are therefore not covered by Article 4's TDM exception. They may, however, qualify for the temporary reproduction exception in Article 5(1) of Directive 2001/29, depending on the circumstances, including lawful access and the use enabled by those copies.

The AG identifies two routes by which retrieval and reuse may trigger the publisher's reproduction right. One concerns access: the temporary copy exception may not apply where a service circumvents effective technical measures restricting access or scraping. The other concerns the response: even if access is lawful, an answer may reproduce a publication if its protected form of expression remains recognisable and the answer allows the user to dispense entirely with the original.

This distinction is relevant because lawful access does not authorise every subsequent use. Likewise, not every chatbot answer drawing on a news article reproduces protected expression. The press publishers' right does not, and should not, confer a monopoly over news as information. A chatbot can report the same facts without reproducing a publication's protected form, even if its answer reduces traffic to the publisher's website. Conversely, recognisability may extend beyond verbatim quotation to elements such as arrangement, structure and phrasing. For making available to the public, AG Szpunar also considers whether the publication's expression is systematically recognisable in responses to related queries (paras 193-216).

This approach preserves the core copyright law distinction between information and expression while acknowledging that a generative AI system can reproduce protected content without simply copying and pasting. Still, the proposed requirement that users be able to dispense *entirely* with the original is demanding. In my view, the Court should clarify how substitution relates to recognisability and how both are to be assessed across changing prompts and responses. This is far from an easy task.

Provider responsibility raises an additional question. AG Szpunar considers that relevant acts may be attributed to the chatbot provider

where they are inherent in its business model or where it fails to take measures expected of a 'reasonably diligent operator'. The fact that a user initiates a query does not automatically determine the issue. Much will depend on what safeguards are reasonably expected and how courts distinguish occasional failures from features of a service's design.

Applying his framework provisionally, the AG suggests that the conditions for liability do not appear to be met on the example before him. The final factual assessment remains for the national court (para. 239).

Beyond *Like Company*

The Opinion will influence discussion of two further Hungarian references concerning generative AI and copyright: *Manchester United Fanatics Klub* (C-806/26) and *Mobilguru* (C-829/26). If the Court follows AG Szpunar on admissibility in *Like Company*, later cases may offer a more appropriate setting to decide issues concerning training that he addresses only in the alternative.

The same caution applies to the cited German litigation. The Opinion is not binding, and the press publishers' right invoked in *Like Company* is distinct from authors' rights. Nevertheless, its analysis of training copies, TDM and memorisation, with which I largely agree, provides a useful framework against which national courts' reasoning can be tested, as early as the pending appeals in those cases.

For EU policymakers, the lesson is that copyright reform must address several different problems with precision. Territoriality and the practical operation of rights reservations merit attention. So do the relationship between retrieval and existing exceptions, and the safeguards expected of providers. But copyright alone cannot resolve concerns about dis/misinformation, source diversity or the economic effects of AI-generated news. It is a fantasy to pretend that it can.

For *Like Company*, the Court should resist turning an uncertain factual record into a general ruling on AI training. It should nevertheless provide clear guidance on retrieval and outputs, preserving the distinction between protected expression and information. AG Szpunar's Opinion offers a thoughtful starting point, and a reminder

that the complexity of the AI landscape becomes more manageable when its legal questions are carefully distinguished and nuanced.

João Pedro Quintais Quintais is Associate Professor at the Institute for Information Law (IViR), University of Amsterdam, and member of the European Copyright Society. Funding acknowledgement: João Pedro Quintais's research was supported by the Dutch Research Council (NWO) through the Vidi project Generative AI Content Moderation: Regulation for Fundamental Rights (file number VI.Vidi.241R.016; grant DOI: 10.61686/HTKXT49425)