The copyright question **Can AI learn form anything and generate whatever it wants?**

Summary

Can AI freely use content available online to learn? The request for a preliminary ruling in Mobilguru (C-829/26) brings the limits of text and data mining (TDM) back into focus. We examine recent European court decisions and the questions referred to the Court of Justice of the European Union concerning the scope of TDM, the distinction between AI training and content generation, the interplay with third-party rights, and the requirements for rights holders to reserve their rights effectively.

BY

Mª Elena Juan Serna

Baylos IP abion

Generative artificial intelligence (AI) has transformed the way we create, work and interact with technology. It can compose songs, design marketing campaigns and generate hyper-realistic videos. It also makes everyday tasks easier, such as findinginformation, and can even create a caricature of us based on everything it knows – or thinks it knows – about us.

To make all this possible, AI needs to learn – and learn a great deal. It does so by analysing and processing vast amounts of information. While lawmakers and courts are seeking to balance technological innovation with the protection of intellectual propertyrights, fundamental questions remain: **Where do the limits lie? Can AI learn from any source and generate any kind of content?**

The text and data mining (TDM) exceptions under Directive (EU) 2019/790 allow certain uses of protected content, subject to legal conditions.

On paper, the framework seems clear. Directive (EU) 2019/790 on copyright in the Digital Single Market (DSM) sought to harmonise practices relating to *text* and *data mining* (TDM), a technique for analysing large volumes of information and content that is essential to training AI systems. The Directive introduced two exceptions allowingcontent to be reproduced and extracted without the authorisation of intellectual property rights holders, provided that the legal requirements are met (referred to below collectively as the **'TDM exceptions**'). The TDM exceptions covers uses for scientific research by organisations and cultural heritage institutions (Article 3) and a broader range of uses, including for commercial purposes (Article 4), subject to the applicable legal requirements, including: (i) lawful access to the content in both cases; and (ii) for uses covered by Article 4, the additional requirement that rights holders have not expressly opposed such use in an appropriate manner (an 'opt-out').

In practice, however, the TDM exception raises questions of legal certainty. In July 2026, a Hungarian court submitted a request for a preliminary ruling to the Court of Justice of the European Union (CJEU) in Case C-829/26. The reference arose from a dispute between Hungarian press publisher Mobilguru and Google Ireland over the latter alleged unauthorised use of Mobilguru's editorial content in generative AI services. The referring court asks **11 questions (see here) covering among others, the concept of an 'information society service', the scope of press publishers' rights under Article 15 of DSM and the limits of the TDM exceptions, , the interplay between these issues, and the requirements for an effective reservation of rights and the appropriate way to reserve rights.** We highlight three key issues raised by the Hungarian court below:

- Does an 'information society service' cover only AI responses, or also model training by the same provider?
- How do the text and data mining exceptions in Articles 3 and 4 of Directive 2019/790 interact with press publishers' rights? Should a distinction be drawnbetween operations carried out during AI training and those involved ingenerating responses?
- Must rights holders use machine-readable technical mechanisms, such as arobots.txt file, or could an express prohibition in a website's terms and conditions, or one communicated directly to the AI developer, suffice?

Across the Atlantic, related questions about the scope of *data mining* and use of protected content have also arisen although under a different legal framework notably the "fair use" doctrine. On 27 December 2023, The New York Times brought proceedings against OpenAI and Microsoft (Case No. 1:23-cv-11195) over the alleged unauthorised use of its news articles to train their AI models.

Rights holders need to understand when and how they can reserve their rights to prevent the use of their content for TDM.

In Europe, German courts have already addressed some of these issues. In Robert Kneschke v LAION, the Hamburg Higher Regional Court ruled on 10 December 2025 that LAION's downloading and analysis of a photograph to build a dataset for AI training were

covered by the German TDM exceptions implementing Articles 3 and 4 of the DSM Directive. First, the Court consider that the claimant had not shown that the opt-out made in the website terms was machine-readable when the photograph was downloaded in 2021. Second, the Court considered dataset creation to be scientific research: possible later use by commercial AI developers did not, in itself, excludeLAION to benefit from the research exception (not affected by an opt-out).

The judgment is not final: an appeal has been lodged with the German Federal Supreme Court.

In GEMA v OpenAI, the Munich Regional Court substantially upheld the collecting society GEMA's claim in its judgment of 11 November 2025. It found that OpenAI's models had stored copyright-protected song lyrics from GEMA's repertoire and could reproduce them in responses to ChatGPT users. The judgment highlights the need to distinguish analytical uses of protected material from the memorisation of works within a model and their reproduction in outputs. According to the court, the TDM exception did not cover the memorisation and outputs at issue. The Judgment has been appealed and is therefore not final.

Defining the limits of TDM in AI training and content generation is central to balancing innovation and copyright protection.

Related questions concerning the distinction highlighted in GEMA between the use of protected content within a model and its reproduction in responses, have also arisen at EU level. In his Opinion of 3 November 2026 in Like Company (C-250/25), Advocate General Szpunar addresses a dispute between a Hungarian operator of news portals, concerning the alleged unauthorised use of news articles by Google's AI chatbot, Gemini, to generate responses to users' queries. He considers that retrieving and reusing press publications to generate chatbot responses may, in certain circumstances, infringe publishers' rights.

Against this backdrop, the Hungarian court's questions in Mobilguru case – and, above all, the CJEU's answers – will be of particular interest to practitioners, AI developersand intellectual property rights holders. They should help clarify how the TDM exception applies at the different stages of AI operation, where its limits lie and how rights holders can exercise their rights.

We will be following developments in Luxembourg closely.