

Kluwer Copyright Blog

Do AI models dream of dolphins in lake Balaton?

Paul Keller (Institute for Information Law (IViR)) · Wednesday, May 28th, 2025

There is a bit of excitement in copyright circles about the first case referred to the CJEU that directly addresses the intersection of artificial intelligence (AI) and the EU copyright framework. The request for a preliminary ruling — [Like Company v Google \(C-250/25\)](#) — originates from the Budapest Capital Regional Court (Budapest Környéki Törvényszék) and involves a dispute between Like Company, a publisher and operator of various online news portals, and Google, in its capacity as the operator of the Bard (now Gemini) chatbot.



ChatGPT based on the input of millions of unknown creators of visual artworks on the public internet

Like Company claims that responses provided by Bard, in reply to requests to summarize the content of a specific web page, infringe its rights under the relevant national and EU legislation (copyright and/or the neighbouring right for press publishers), as the response constitutes an unauthorized communication to the public. Whether chatbot answers that summarize publicly available information protected by the press publishers' right constitute a communication to the public indeed seems like an interesting new question for the CJEU to answer^[1] — and one I'll gladly leave to [more qualified people to opine on](#).

Instead, I will focus on another — somewhat problematic — aspect of the referral: it appears to misrepresent some of the underlying technical processes, which has led the court ([and some commentators](#)) to frame the central issue as one concerning the legality of training AI models on publicly available content. In the second and third questions referred to the CJEU, the Budapest Capital Regional Court asks (emphasis mine):

- *Must Article 15(1) of Directive 2019/790 and Article 2 of Directive 2001/29 be interpreted as meaning **that the process of training an LLM-based chatbot** constitutes an instance of reproduction, where that LLM is built on the basis of the observation and matching of patterns, making it possible for the model to learn to recognise linguistic patterns?*
- *If the answer to the second question referred is in the affirmative, does such reproduction of lawfully accessible works fall within the exception provided for in Article 4 of Directive 2019/790, which ensures free use for the purposes of text and data mining?*

And while the latter question is indeed the billion-euro question when it comes to the applicability of the EU copyright framework to AI training — and one that the CJEU will likely have to answer at some point — the connection between this issue and the facts at hand in *Like Company v Google* seems spurious at best. Yes, there is little doubt that Bard (now Gemini) is based on an AI model trained on large amounts of copyright-protected (and non-protected) material sourced from the public internet. But based on the facts as established by the Budapest District Court, it seems highly improbable that the alleged infringement results from reproductions made during the training of the AI model that the chatbot in question was based on.

The underlying facts that gave rise to the dispute are presented in points 7 and 8 of the “succinct presentation of the facts and procedure in the main proceedings” section of the referral document:

1. *An article appeared on one of the applicant’s protected online press publications ([balatonkornyeke.hu](https://www.balatonkornyeke.hu)) stating that Kozsó, a well-known Hungarian singer, had not given up on his dream of putting dolphins in an aquarium next to Hungary’s largest lake, Lake Balaton. That article also made reference to other online press publications belonging to the applicant, reporting on the hospitalisation of Kozsó, his interests, the fact that he had served a custodial sentence in the United States and also a fine he had received for electricity theft.*
2. *In response to the question ‘Can you provide a summary in Hungarian of the online press publication that appeared on [balatonkornyeke.hu](https://www.balatonkornyeke.hu) regarding Kozsó’s plan to introduce dolphins into the lake?’, the defendant’s chatbot provided a detailed response which included a summary of the information appearing in the news media belonging to the applicant.*

Dolphins in Lake Balaton

The description in point 7 makes it very likely that the article at issue is [Kozso nem adja fel: továbbra is delfinek szeretne a Balatonhoz telepíteni a népszerű énekes](#) (which translates to “Kozso doesn’t give up: the popular singer still wants to introduce dolphins to Lake Balaton”), published on 21 July 2023.

It is the description of the actual mechanics of the case in point 8 that makes it clear this case is not about the training of AI models, but about something else entirely. What seems to have occurred is that a user — with prior knowledge of the article in question — directed the chatbot to provide a summary by referencing the domain name of the publication where the article was published and providing enough contextual information to identify the specific article. In response, the chatbot (an LLM) accessed the content of the website and generated a summary of the text found there.

Given the close temporal proximity between the publication of the article (21 July 2023) and the period for which infringement is alleged (13 June 2023 to 7 February 2024), it seems highly unlikely that the underlying model had been trained on the content of that specific article^{[2],[3]}. Instead, it appears almost certain that the already trained model used the live content of the website as input, and then operated on it to produce the requested summary. This interpretation is also supported by the defendant’s explanation, summarized in point 23: “*In order to collect data, [the chatbot] uses the Google Search database, and, in its response, it is able to display a modified version of an article, if the user has already provided the original version of the article in his or her instructions.*” In other words, upon receiving the prompt, the chatbot searched the Google Search index for content from the referenced website and then produced a summary based on that

content – a type of process often referred to as Retrieval Augmented Generation (RAG).

While such interactions with chatbots — and their ability to summarize websites on demand — may still seem novel, the overall process is not. Attentive readers may notice that the translation of the article provided above via Google Translate is the result of an analogous process. Given a pointer to the article (in this case, the URL), a service operated by Google (Google Translate) uses the content of the website as input for an AI model, which then transforms it into the requested output (an English translation). The only substantive difference is that, in the translation case, Google goes to great lengths to preserve the overall structure and context of the original website^[4], whereas in the summary case, the output is presented within the chatbot interface, which bears little or no relation to the source website.

Based on all of this it seems safe to conclude that the case as referred to the CJEU does not in fact deal with issues related to the training of AI models but rather with issues arising from their use. This distinction is important for at least two reasons: On a practical level there is a real danger of arriving at conclusions that can limit the freedom of individual users to interact with publicly available content based on mistaken understanding of the underlying technology. And on a more general level it seems important that decisions related to the applicability of the TDM exception to AI training will be made based on a case that actually involves AI training. As I have shown above that is almost certainly not the case here at least not in the terms described by the court.

^[1] [The article in question](#) at the center of the dispute certainly makes a great addition to the eclectic CJEU case law on communication to the public.

^[2] Training large AI models such as bard generally takes months and they commonly have knowledge-cut off dates that are well before they are deployed.

^[3] Note that there is a slight inconsistency here between the publication date and the presentation of facts that alleges that the making available to the public occurred between 13 June 2023 and 7 February 2024. The most likely explanation is that one of the dates is not correct.

^[4] This includes the provision of a URL that makes great efforts to appear as if the content is hosted on the original website, but that at closer inspection reveals itself as a URL fully controlled by Google: translate.google.com

To make sure you do not miss out on regular updates from the Kluwer Copyright Blog, please subscribe [here](#).

2024 Future Ready Lawyer Survey Report

Legal innovation: Seizing the future or falling behind?

[Download your free copy →](#)



Wolters Kluwer



This entry was posted on Wednesday, May 28th, 2025 at 1:07 pm and is filed under [Artificial Intelligence \(AI\)](#), inter alia, for ensuring that EU law is interpreted and applied in a consistent way in all EU countries. If a national court is in doubt about the interpretation or validity of an EU law, it can ask the Court for clarification. The same mechanism can be used to determine whether a national law or practice is compatible with EU law. The CJEU also resolves legal disputes between national governments and EU institutions, and can take action against EU institutions on behalf of individuals, companies or organisations.”>[CJEU, European Union, Text and Data Mining \(TDM\)](#)
You can follow any responses to this entry through the [Comments \(RSS\)](#) feed. You can leave a response, or [trackback](#) from your own site.