GT | GreenbergTraurig

Alert | Innovation & Artificial Intelligence / Intellectual Property & Technology

EU AI Act's Opt-Out Trend May Limit Data Use for Training AI Models

With the introduction of the EU AI Act, companies that train AI models should be vigilant of which datasets they use. Certain copyright-protected works may no longer be available for use in training AI models.

Introduction

There are three primary aspects of generative AI applications that intersect with copyright protection: (1) machine learning with protected works (*the input side*); (2) the protectability of works created with the assistance of generative AI (*the discussion on copyright protection of AI generated works*); and (3) potential infringement by the output of preexisting works (*the output side*). This GT Alert focuses on legal issues with respect to the input side.

On the input side, a main concern is whether publicly available, copyright-protected works may be used to train AI models used by commercial companies.

Because AI is designed to mirror human intellect, the input side of AI may be compared to a person reading a book or listening to music. The process of gaining knowledge or inspiration through reading the book or listening to music is seamless and happens automatically. There are no laws prohibiting the use of the increased knowledge or inspiration for commercial purposes, barring any direct or indirect copying from the respective books or music.

Compared to the organic way that humans gain knowledge, those who train AI models have more control over how and whether AI learns from data fed to it. Another difference between AI and human intelligence is AI's capacity to consume and build on ostensibly unlimited quantities of data. As a result, AI has the potential to accelerate technological processes and innovation exponentially.

Consequently, an emerging question surrounding AI regulation is the extent to which machine learning should be limited in order to respect intellectual property rights.

Machine Learning under the EU AI Act

The EU AI Act, approved by the EU Council 21 May 2024, takes a first stab at answering this question. The EU AI Act contains a provision that equates AI/machine learning with "text-and-datamining" (TDM) under the EU Text and Data Mining Directive.[1] Consequently, "machine learning" is allowed, provided that

- the person programming the machine-learning functionality has had **lawful access** to the content for the purpose of text and data extraction; and

- the owner of the copyright and related rights and/or the database owner have **not expressly reserved the extraction** of text and data (the so-called opt-out mechanism).

The EU AI Act is expected to enter into force in 2024 and will fully apply 24 months thereafter. However, the TDM exception under the EU Text and Data Mining Directive already exists. Therefore, the TDM exception for machine learning can already be enforced in anticipation of the EU AI Act's interpretation.

Opt-Out Mechanism

While it is not yet clear how a legally valid opt-out request would be made, various organizations, such as the Dutch collecting society Pictoright (pictures) and the French collecting society Sacem (music), have drafted general reservation of rights statements, allowing creators to opt out of their data being used to train AI models. In addition, many websites and images on social media now include similar opt-out statements.

While no case law or other authoritative document exists to determine the sufficiency of such statements to trigger the opt-out criteria, the trend may intensify now that the EU AI Act has been adopted.

Takeaways

Although the EU AI Act and its TDM exception have yet to become formally applicable, AI system providers and developers should consider implementing measures and configurations to avoid infringement claims by rightsholders. Below are four additional considerations.

- **Obtain legitimate access to the content**. In the process of web-scraping or reviewing pre-built datasets, confirm whether the content used for machine learning purposes is subject to access restrictions, such as a paywall or other (technical) restrictions.

---

[1] Article 52c(1) of the EU AI Act; and Article 4 EU Text and Data Mining Directive.

- **Verify the Existence of Opt-Out Reservations**. Consider verifying that the rightsholders have not reserved the right to make reproductions for TDM purposes by, for example, searching collective rights organizations' websites to see which works include the opt-out criteria.

- **Include the Necessary Contractual Protections**. Two types of agreements are of particular relevance in terms of machine learning: (1) agreements with the owner of the dataset on which the AI model is trained; and (2) agreements with the customer of the AI model. In both cases, consider creating an agreement with a fair and balanced distribution of liability regarding inadvertent use of opted-out copyright-protected works.

- **Place Guardrails on the AI Model to Prevent Use for Other Purposes Beyond TDM**. For copyright-protected content that may be used for TDM purposes, consider implementing technical and organizational limitations on the use of the content to ensure it is only used for training the AI model.

Authors

This GT Alert was prepared by:

- Wouter van Wengen | +31 20 301 7445 | Wouter.vanWengen@gtlaw.com

- Radboud Ribbert | +31 20 301 7333 | Radboud.Ribbert@gtlaw.com