

# AI and Fair Use: A Discussion in Comparative Law

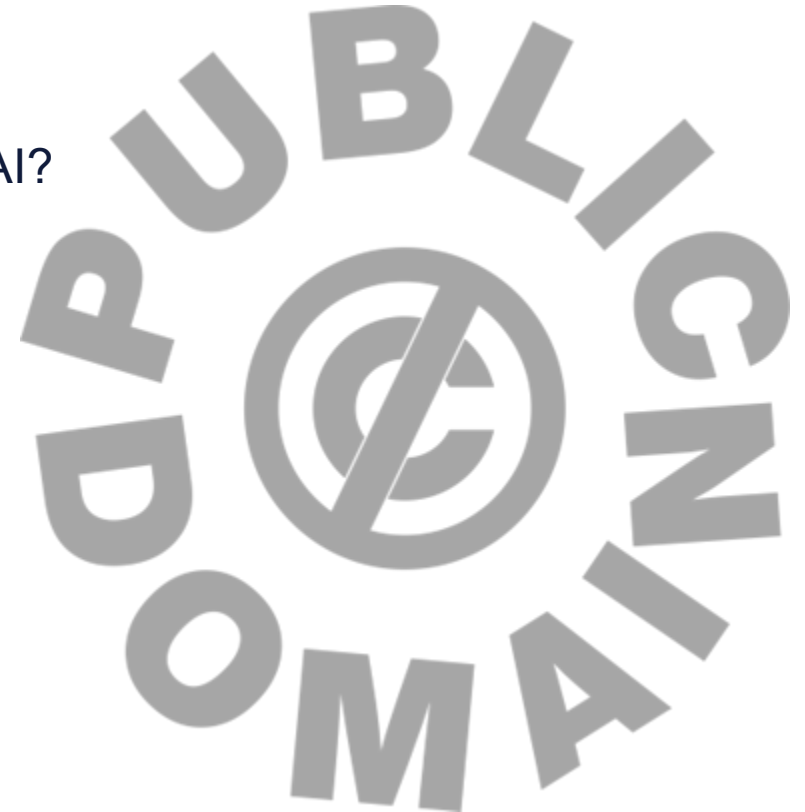
ASIPI 2025, Wednesday, 3 December, Buenos Aires, Argentina  
Mladen Vukmir, Founding Partner  
VUKMIR + ASSOCIATES LLC  
Zagreb, Croatia

# Overview

- + **AI and Copyright: Open questions**
- + EU legal framework
- + European case law, studies and reports
- + **Future regulation of AI & Copyright**
- + Ensuring transparency – is it possible?
- + **An alternative regulation model**
- + What is creativity?
- + Limitations of applying current copyright frameworks to AI technologies
- + **So?**

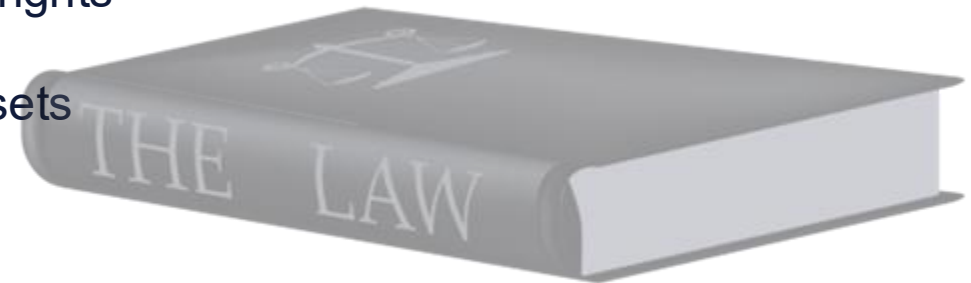
# AI and Copyright: Open questions

- + Is AI diminishing, bolstering, or simply changing human **creativity**?
  - Can AI create at all – could it?
  - How many humans can create on the same level as the AI?
- + Limits of fair use and exceptions applied to AI training
- + Fair compensation to authors
- + Prior authorization & opt-out – a solution or a problem?
- + Ensuring transparency in AI training datasets – is it possible?
- + Legal uncertainty over ownership of AI-generated content



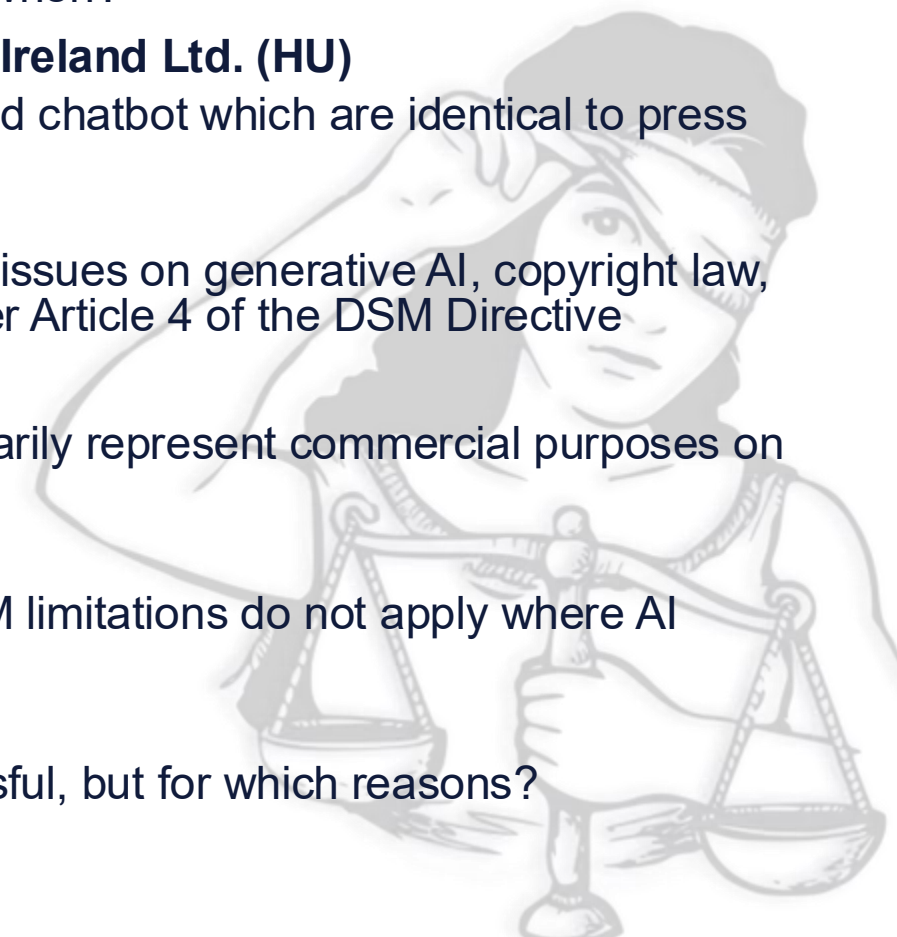
# EU Legal Framework

- + Is the EU regulation an attempt to answer the questions we have mentioned?
- + InfoSoc Directive, AI Act and Directive on Copyright in the Digital Single Market (CDSM)
- + AI Act – Obligation to publish summaries detailing training data and copyright status
- + CDSM – text and data mining (TDM) exceptions
- + Rights holders opt-out option from data mining via reserved rights
- + AI providers must respect opt-out mechanisms limiting datasets



# European case law

- + Cases might drive evolution towards more coherent and consistent AI copyright frameworks. Will they? In which direction will national courts go? Will EU court(s) agree and when?
- + **Court of Justice of EU case C-250/25 Like Company v. Google Ireland Ltd. (HU)**
  - CJEU asked to clarify whether the responses of an LLM trained chatbot which are identical to press articles constitute an act of reproduction
  - the decision not expected before 2027
  - the case is closely watched across the EU as it will clarify key issues on generative AI, copyright law, and the application of the text and data mining exception under Article 4 of the DSM Directive
- + **Germany - Hamburg District Court, LAION v. Robert Kneschke**
  - Web scraping for compiling training datasets does not necessarily represent commercial purposes on its own
- + **Germany - Munich Regional Court, GEMA v. OpenAI**
  - Processes within LLMs can be classified as reproduction; TDM limitations do not apply where AI systems are able to reproduce their training data
- + **England & Wales – Getty Images v. Stability**
  - Getty Images' copyright infringement claims proves unsuccessful, but for which reasons?



# More recent European case law

- + **Germany - Munich Regional Court: GEMA v. Sunny** – hearing was scheduled for the November 26
  - alleges unauthorized use of song lyrics in training an AI model that then generates too similar or identical text, which means a violation of copyright.
  - The regional court decides on individual cases - interpretation at the level of the European Court of Justice in relation to EU law.
- + **Denmark - Copenhagen City Court: KODA v. Suno (USA)** – filed on November 4, 2025.
  - A lawsuit for the use of copyrighted works of KODA members (a Danish collective organization) without permission, remuneration and transparency
  - KODA claims that Suno uses the protected works of its members without permission to train the model and then generates works that are very similar (or identical) to these works.
  - It uses the technique of "stream-ripping" from the YouTube platform and scraping texts without the permission of the right holder



# EU Studies and Reports

- + **Axel Voss (MEP)** new compromise proposals to JURI Report on Copyright and GenAI
- + **Danish Presidency of the Council of the EU 2025** draft summary report on GenAI Licensing
- + **EP Generative AI and Copyright study by Nicola LUCCHI, PhD**
  - commissioned by the **EP Committee on Legal Affairs**, analyses how GenAI challenges the fundamental principles of EU copyright and points out the legal inconsistency between the practice of training the AI model and the existing exceptions for text and data mining, as well as the unclear status of the content generated by AI.
  - connects Art. 53 of the AI Act, with Recital 105. which "presuppose" the use of the author's exclusive rights which "can be widely used" in the context of training AI models, arguing that these provisions do not have the material effect which would extend copyright exceptions.
  - emphasizes that Art. 53 of the AI Act has a procedural nature - it imposes transparency and compliance obligations on providers of general-purpose models but does not grant new rights or expand the material scope of copyright exceptions in the EU law. Moreover, the obligations of Art. 53 only apply if an exception for TDM can at all be made legitimately applied - **meaning it does not provide any automatic "shield" to AI platforms.**
  - It points out that **it is crucial to develop a licensing model that enables transparency and fair distribution, compensation and enable effective negotiations with AI services**

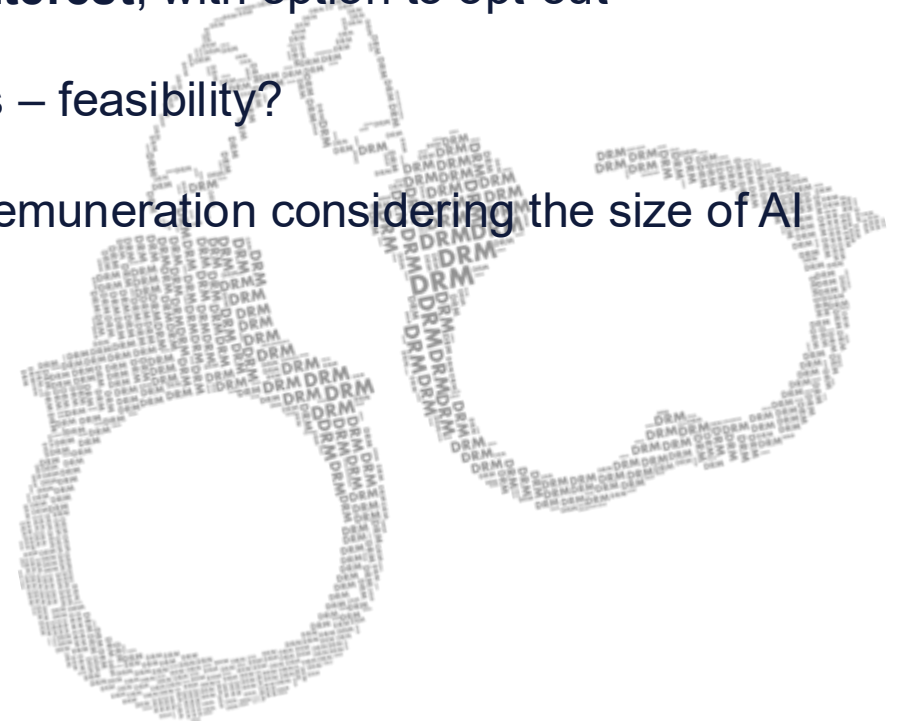
# EUIPO Copyright conference Alicante 20.-21. November

## + A participant observations (Burak Ozgen):

- For creators' community there is no way but asking for additional measures at EU level to develop a sustainable GenAI market that properly remunerates creators
- Big Tech/AI still continues to ignore the pain points, and EU tech/AI sector struggles with anti-competitive disadvantages due to big tech's refusal to comply with EU rules
- There is a growing understanding and willingness among the EU and national policy makers to improve the EU legal framework to create a functioning licensing market and to better enforce EU rules to all players in the market, which is encouraging
- What was heard several times this week: "2026 will be an important year for copyright in the EU"

# Future regulation of AI & Copyright

- + AIPPI AI & Copyright study – Yokohama discussions
- + Prior authorization from copyright holders as a principle based on the exclusive right – feasibility?
- + Exceptions authorizing TDM of copyrighted works **for public interest**, with option to opt-out
- + Imposing transparency obligations on developers of AI systems – feasibility?
- + Brazil's AI Bill 2,338/2023 – adopting the opt-out system, with remuneration considering the size of AI companies and frequency of use of protected materials
- + Australia – stepping away from TDM exceptions



# Ensuring transparency – is it possible?

- + Providers do not disclose AI training data sources transparently
- + Technical measures (robots.txt) protect copyrighted material from crawlers only to a limited degree
- + A voluntary model proposal:
  - Collaboration between creators, AI developers, and policymakers
  - Encouragement of balanced policies supporting creators and AI innovation
  - Creation of an EU-wide registry of rights and AI data licenses



Join Wikipedia Asian Month this November and December!  
 Contribute in Wikipedia Asian Month and get a postcard!  
[\[help with translations\]](#)

# robots.txt

27 languages

Contents

(Top)

[History](#)

[Standard](#)

Compliance

[Search engines](#)

[Archival sites](#)

[Artificial intelligence](#)

[Security](#)

[Alternatives](#)

Examples

[The use of the wildcard \\* in rules](#)

Nonstandard extensions

[Crawl-delay directive](#)

[Sitemap](#)

[Universal \\*\\*\\* match](#)

[Content-Signal](#)

Meta tags and headers

[A "noindex" meta tag](#)

[A "noindex" HTTP response header](#)

[Maximum size of a robots.txt file](#)

[See also](#)

[References](#)

[Further reading](#)

[External links](#)

Article [Talk](#)

[Read](#) [Edit](#) [View history](#) [Tools](#)

From Wikipedia, the free encyclopedia

Checked

**robots.txt** is the filename used for implementing the **Robots Exclusion Protocol**, a standard used by websites to indicate to visiting web crawlers and other web robots which portions of the website they are allowed to visit.

The standard, developed in 1994, relies on *voluntary compliance*. Malicious bots can use the file as a directory of which pages to visit, though standards bodies discourage countering this with *security through obscurity*. Some archival sites ignore robots.txt. The standard was used in the 1990s to mitigate *server* overload. In the 2020s, websites began denying bots that collect information for *generative artificial intelligence*.

The "robots.txt" file can be used in conjunction with *sitemaps*, another robot inclusion standard for websites.

Search engines use crawlers (bots) to index website content. Without guidance, these bots may crawl unnecessary or irrelevant pages. The Robots.txt file helps control what search engines should or should not index.<sup>[1]</sup>

## History [ edit ]

The standard was proposed by [Martijn Koster](#),<sup>[2][3]</sup> when working for [Nexor](#)<sup>[4]</sup> in February 1994<sup>[5]</sup> on the *www-talk* mailing list, the main communication channel for WWW-related activities at the time. [Charles Stross](#) claims to have provoked Koster to suggest robots.txt, after he wrote a badly behaved web crawler that inadvertently caused a *denial-of-service attack* on Koster's server.<sup>[6]</sup>

The standard, initially *RobotsNotWanted.txt*, allowed *web developers* to specify which bots should not access their website or which pages bots should not access. The internet was small enough in 1994 to maintain a complete list of all bots; *server* overload was a primary concern. By June 1994 it had become a *de facto standard*;<sup>[7]</sup> most complied, including those operated by search engines such as *WebCrawler*, *Lycos*, and *AltaVista*.<sup>[8]</sup>

**robots.txt**

Robots Exclusion Protocol

```
User-agent: Mallorybot
Disallow: /

User-agent: *
Crawl-delay: 20
Disallow: /secret/
```

Example of a simple robots.txt file, indicating that a user-agent called "Mallorybot" is not allowed to crawl any of the website's pages, and that other user-agents cannot crawl more than one page every 20 seconds, and are not allowed to crawl the "secret" folder

**Status** Proposed Standard

**First published** 1994 published, formally standardized in 2022

**Authors** [Martijn Koster](#) (original author) [Gary Ilyes](#), [Henner Zeller](#), [Lizzi Sassman](#) (IETF contributors)

**Website** [robotstxt.org](#) ℹ, [RFC 9309](#) ℹ

Appearance

Text

- Small
- Standard
- Large

Width

- Standard
- Wide

Color (beta)

- Automatic
- Light
- Dark



# An alternative regulation model – *sui generis* right?

- + Based on the general legal authorization: a blanket training permission system balanced with a lump-sum and/or market-share levy (on the participation side but distribution remains a problem)
- + Broad public interest justification: learning from Japan's approach
- + (Re)Assessing the role of the authors' rights exceptions under the three-step test
- + Utilizing new or existing collective rights management organizations for redistribution of funds to copyright holders
- + Voluntary **registration** of copyrighted works – a prerequisite for exercise of rights in AI age
- + The return of the (non-invasive) DRM?



# What is human creativity?



“An insight-related neural reward signal” study

# Limitations of applying copyright to AI tech creativity

- + Let's first ask ourselves: will copyright start to limit the human creativity?
- + Human-centric structure & moral rights misfit to AI
- + Berne Convention's concept of originality – a relic of the past
- + The very concept of creativity changed since Berne
- + Threshold of authenticity: a new criteria for the threshold of copyright protection
- + Tackling digital identity rights through copyright – Denmark's initiative!
- + Is move to “data” approach (non-proprietary?) more viable for AI training?



# What is originality?

*"Nothing is original. Steal from anywhere that resonates with inspiration or fuels your imagination. Devour old films, new films, music, books, paintings, photographs, poems, dreams, random conversations, architecture, bridges, street signs, trees, clouds, bodies of water, light and shadows. Select only things to steal from that speak directly to your soul. If you do this, your work (and theft) will be authentic. Authenticity is invaluable; originality is non-existent. And don't bother concealing your thievery—celebrate it if you feel like it. In any case, always remember what Jean-Luc Godard said: "It's not where you take things from—it's where you take them to."*

— Jim Jarmusch

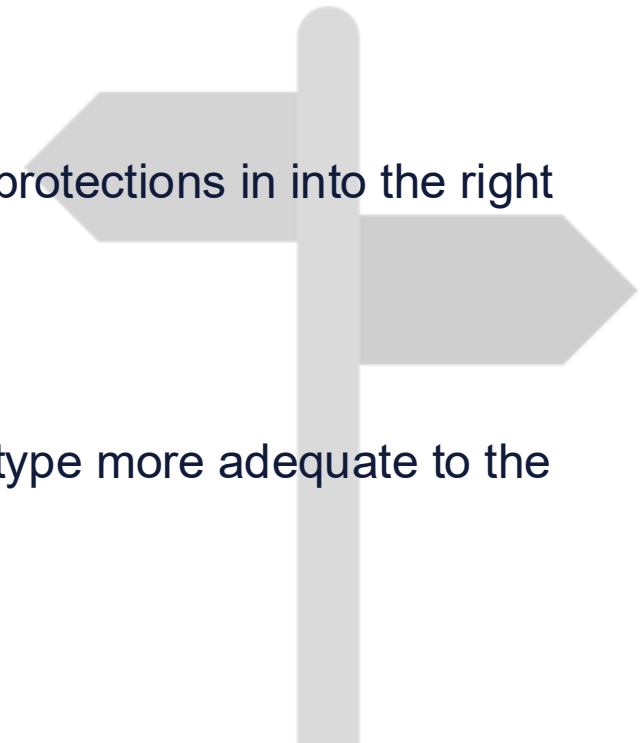
**Authenticity**  
is invaluable.

**Originality**  
is non-existent.



# So?

- + We have already seen a change in creativity from “romantic” gifted individual, inspirational based to the ubiquitous knowledge-based activity of an educated individual.
- + Copyright did not go unscathed in its path becoming an industrial property rights after the entertainment and especially software industries started relying on it for protection
- + It did not survive the digitalization of our realities in its integral form
- + Is the future of copyright in merging of freedom of expression and creativity protections in into the right protecting **authenticity** rather than **originality**?
- + Is the future of enforcement technological rather than judicial?
- + Is the future collection and distribution of compensation of “blank tape levy” type more adequate to the authors’ needs than the one based on the exclusive rights?



# Sources & Further Reading

- + [https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST\\_STU\(2025\)774095\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)774095_EN.pdf)
- + <https://www.euipo.europa.eu/en/publications/genai-from-a-copyright-perspective-2025>
- + <https://aiippi.soutron.net/Portal/Default/en-GB/DownloadImageFile.ashx?objectId=9948&ownerType=0&ownerId=6152>
- + <https://www.theladders.com/career-advice/this-study-explains-why-humans-are-creative>
- + <https://www.sciencedirect.com/science/article/pii/S1053811920302445?via%3Dihub>
- + <https://www.goodreads.com/quotes/131591-nothing-is-original-steal-from-anywhere-that-resonates-with-inspiration>
- + [.....](#)

# Thank you for your attention!

Mladen Vukmir, Founding Partner  
VUKMIR + ASSOCIATES LLC  
Attorneys at law + Patent + Trademark Agents  
Gramača 2L  
10000 Zagreb  
Croatia

[mladen.vukmir@vukmir.net](mailto:mladen.vukmir@vukmir.net)

[www.vukmir.net](http://www.vukmir.net)