

ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW

A GLOBAL OUTLOOK

Dr. Péter Mezei, Professor of Law

University of Szeged, Faculty of Law and Political Sciences (Hungary)

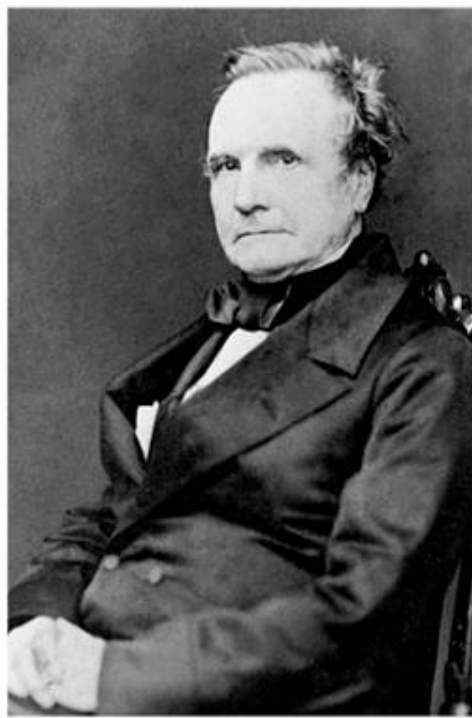
Adjunct Professor (dosentti), University of Turku (Finland)

Professor invité, Université Jean Moulin Lyon III (France)

Chief researcher, Vytautas Magnus University (Lithuania)

Member of the European Copyright Society

mezei.peter@szte.hu



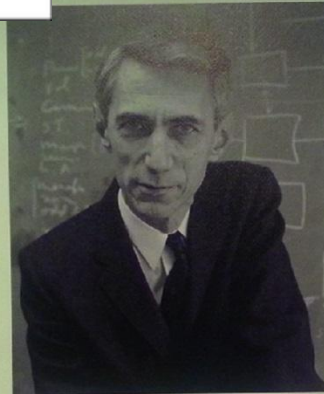
AI & ©

To Claude Shannon and Alan Turing...

[There was no human faculty that couldn't one day be replicated by computers. (Turing)]

From Ada Lovelace and Charles Babbage...

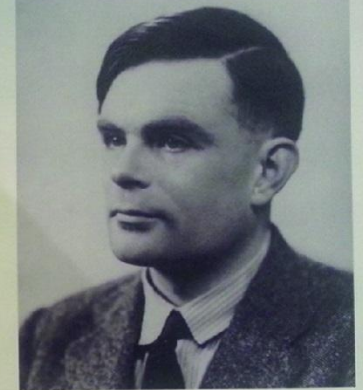
“[The Analytical Engine] has no pretensions whatever to originate anything [and is limited to] whatever we know how to order it to perform” (Lovelace)



Claude Shannon
1916–2001

Claude Shannon was a mathematician and electrical engineer. Whilst working at Bell Laboratories he discussed artificial intelligence and the far-reaching possibilities of digital computers with Alan Turing. For the first time, Shannon gave 'information' a mathematical value. He inspired a new academic field, information theory, which underpinned developments in modern communication.

Image: Getty Images/Science Photo Library



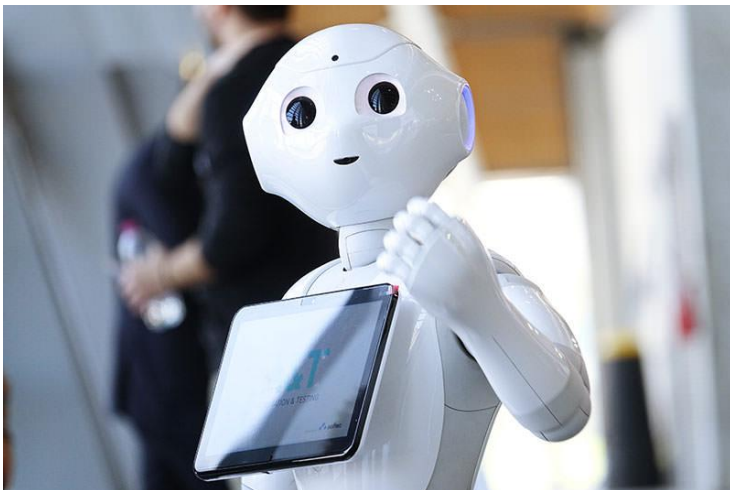
Alan Turing
1912–1954

Alan Turing was a mathematician, philosopher and computing pioneer. With the outbreak of the Second World War, Turing became head of a code-breaking unit at Bletchley Park. He used his mathematical skill to design a series of electromechanical code-breaking machines. His work continued to provide a theoretical basis for post-war developments in computing.

Image: Science Museum/Science Photo Library



"Alexa, play
morning
playlist."



AI IN THE CREATIVE INDUSTRIES

AI in movies



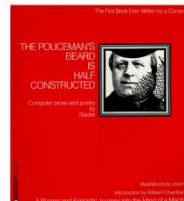
AI in the music industry



AI and arts



AI in literature



AI in the news industry



WHAT IS ARTIFICIAL INTELLIGENCE?

Software - hardware

Strong (full), general or weak (narrow) AI

System, entity, science

“A non-human system that possesses cognitive functions and skills such as learning and reasoning. A smart computer that can think and plan strategically. A science that can assist humanity to find answers to the big questions/themes we face.” (***Kop referring to Hawking***)

“AI is the simulation of human intelligence on a machine, so as to make the machine efficient to identify and use the right piece of ‘Knowledge’ at a given step of solving a problem.” (***Noto La Diega citing Konar***)

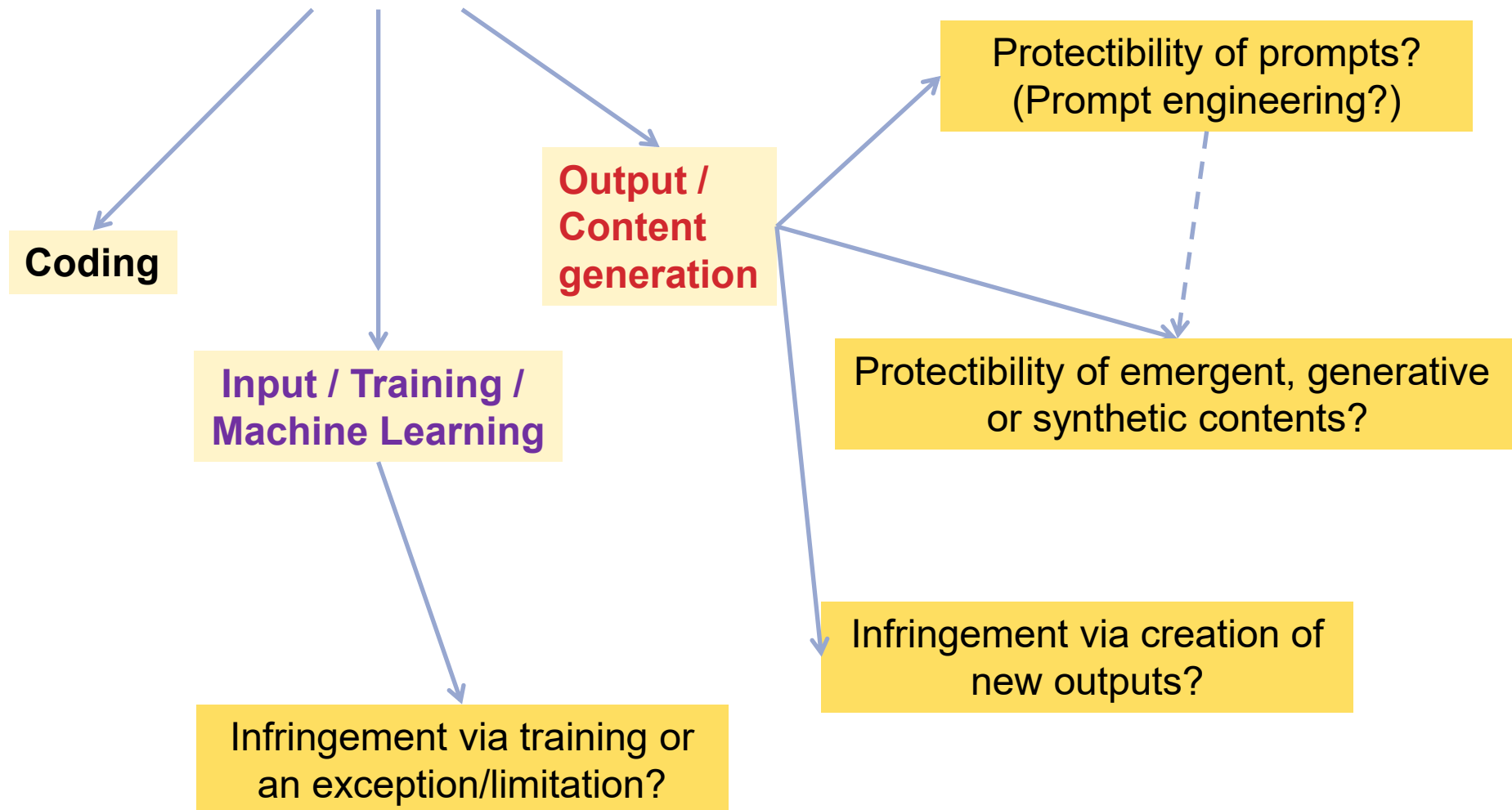
“AI is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment.” (***Nilsson***)

“AI systems will be designed to perform human-like cognitive tasks, steadily improving their performance by learning from experience or external data.” (***De Rouck***)

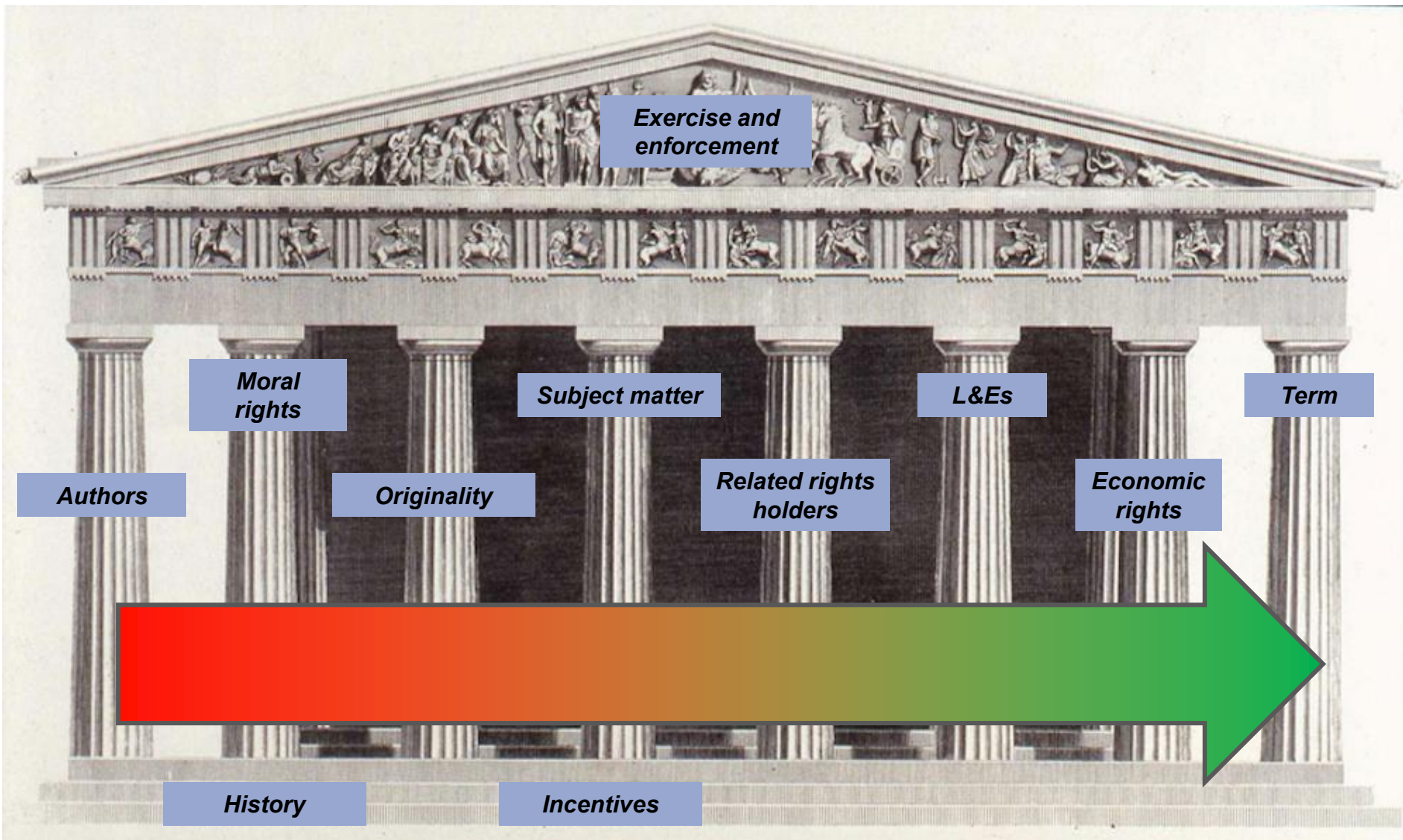
“Defining AI is not an easy task” (***Yanisky-Ravid***)

STAGES OF AI USE

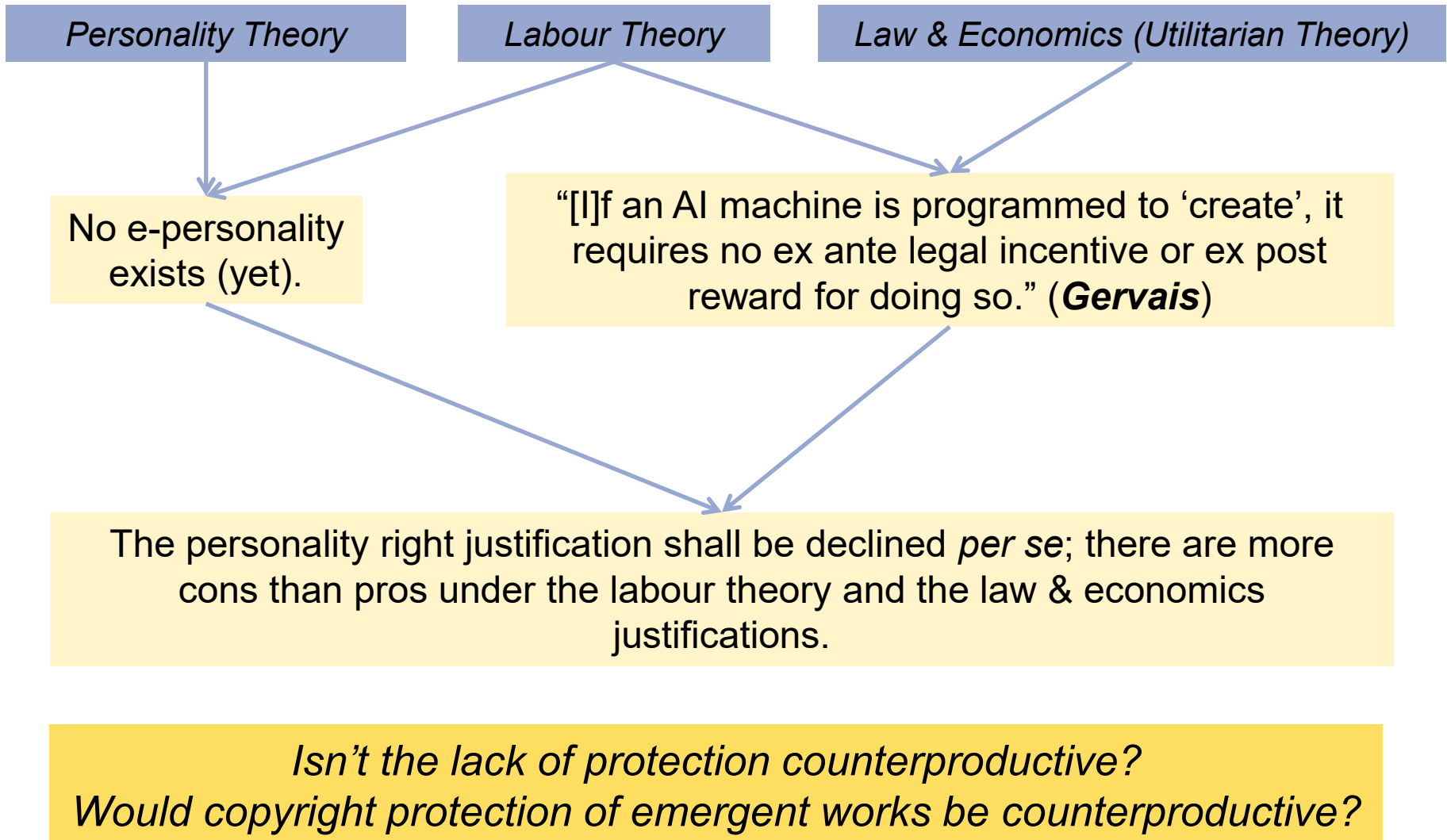
Three main stages of creation of/with AI:



THE RULING ON THE FIELD...



INCENTIVES VERSUS AI



AUTHOR VERSUS AI

Concept of authorship
(originating from the
Berne Convention)

HUMAN
AUTHOR + NATIONAL
OF MEMBER STATES

Domestic laws are based on
this originalist premise

CJEU case law
(*“Author’s own intellectual
creation” + “personal touch”*)

USA: “authorship”
(*“creations of the mind”*)



“Constitutionally, copyright law requires authors; it cannot simply kill them off.”
(**Buccafusco**)

ORIGINALITY VERSUS AI

No international definition
(except reference to “original works” in BC and the idea v. expression dichotomy)

“There [should] be some *intellectual* contribution above and beyond that of simple effort (‘sweat of the brow’)” (**Ricketson**)

CJEU case law
“Author’s own intellectual creation”

USA
“original works of authorship fixed in any tangible medium of expression”

“**Global merger**”

The concept of originality is bound to the concept of authorship; subject matter and creativity.

Compare to closed v. open lists of subject matters!

“Whether a computer can be creative is ultimately a philosophical question” (**De Rouck**)

Copyright is not an investment protection scheme!

MORAL RIGHTS VERSUS AI

Right of paternity

Right of integrity

Right of first publication / Right of withdrawal

The main purpose of moral rights is to build a **strong personal relationship** between the author and her work.

Σ *Moral rights protect the author's "trademarks".*

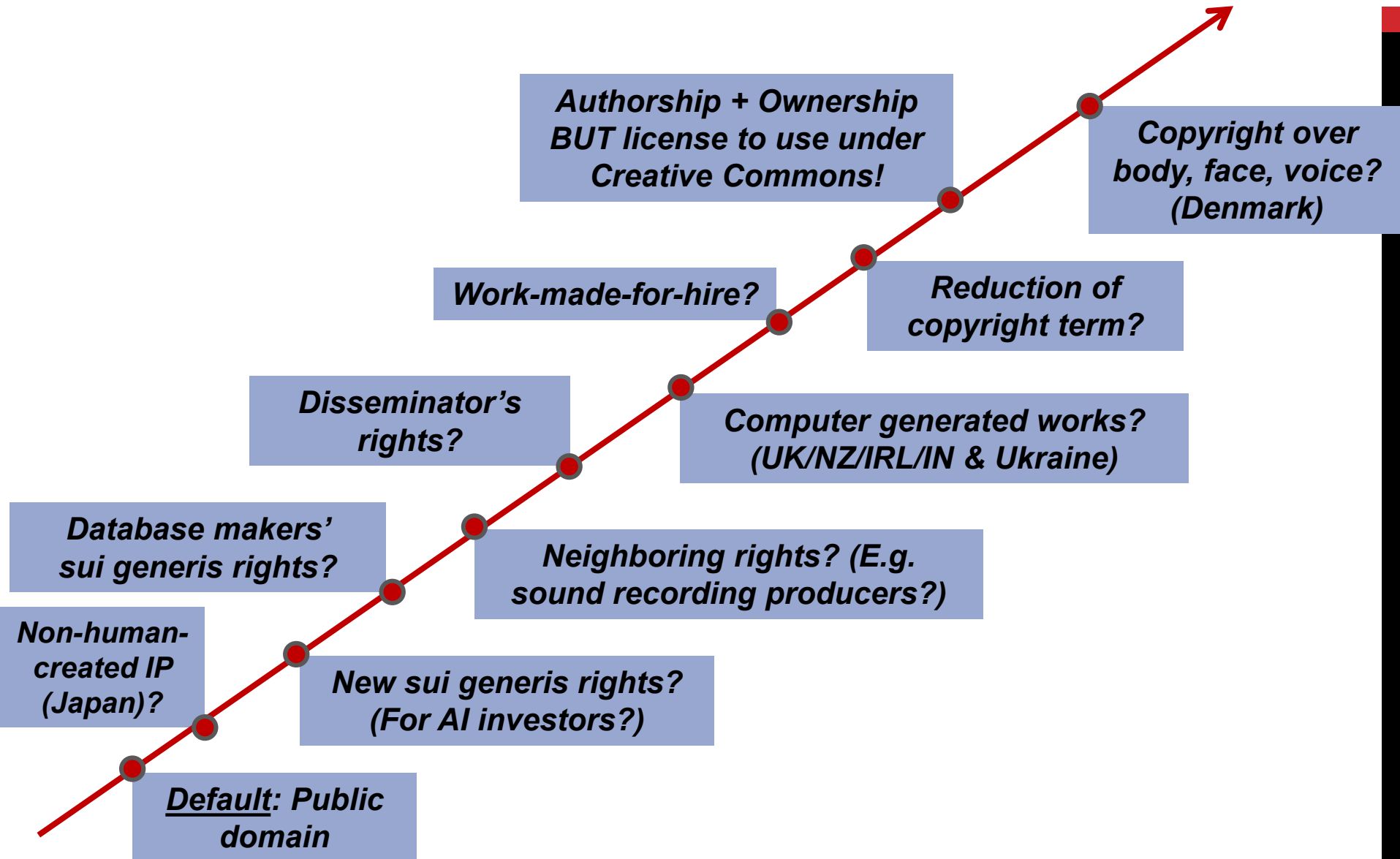
"[T]he link between a work and its author (...) also operates as an indicator of its subject, reliability, and quality." (**De Rouck**)

"[T]he author's name is not simply an element in a discourse (...); it performs a certain role with regard to narrative discourse, assuring a classification function." (**Foucault**)

"Today, the emotional bond between author and work has been loosened". (**Lauber-Rönsberg & Hetmank**)

Can AI "**exercise**" the rights treated to be moral (or personal)?
Can AI have "**interests**" in those (moral) rights?

ALTERNATIVES TO PROTECTION BY COPYRIGHT?



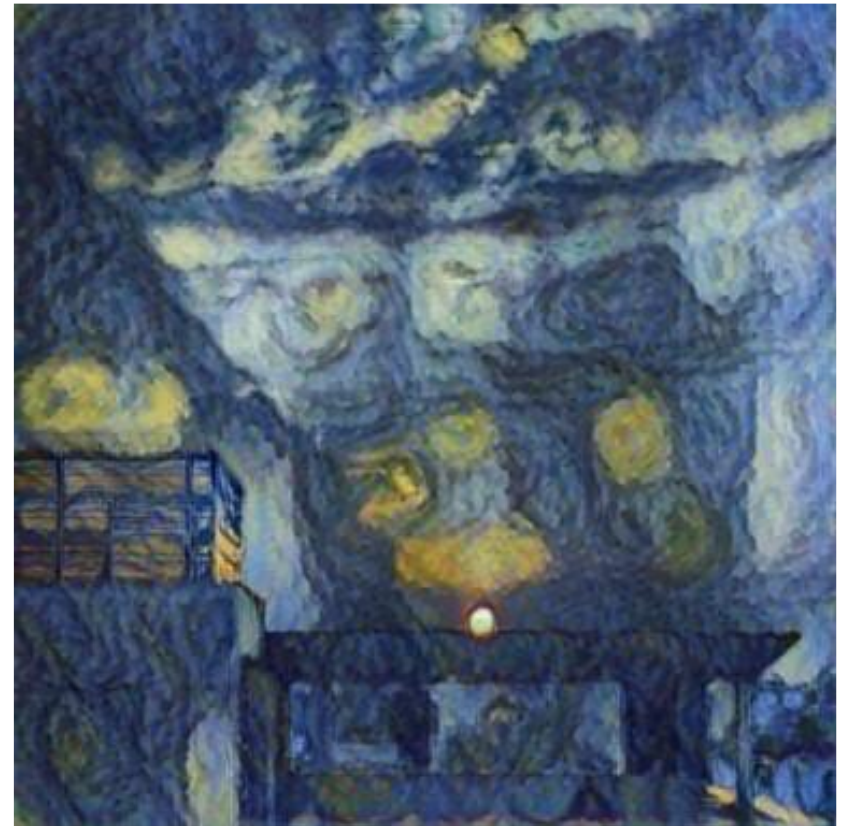
AUTHORSHIP UNDER COPYRIGHT LAW



Q: “Can an AI machine be an author under the copyright Act?”

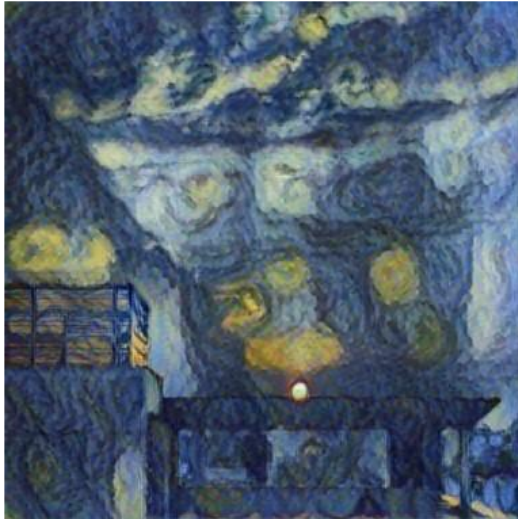
NO.

- The Indian Copyright Act only recognizes works created by **legal entities**
- Author = legal entity = human
- “Modicum of creativity test”
- The work must be “original”
- Regardless, **co-authorship of *Suryast* was initially given to RAGHAV AI Painting App**, but was **withdrawn** on November 25, 2021



Authors:
Raghav Artificial Intelligence Painting App
Ankit Sahni

AUTHORSHIP UNDER COPYRIGHT LAW



Authors:
Raghav Artificial Intelligence Painting App
Ankit Sahni

Canadian Copyright Database

► [Third-Party Information Liability Disclaimer](#)

The links on the text provide access to the help file.

<u>Title:</u>	SURYAST
<u>Type:</u>	Copyright
<u>Registration Number:</u>	1188619
<u>Status:</u>	Registered
<u>Registered:</u>	2021-12-01
<u>Category of Work:</u>	Artistic
<u>Date Published:</u>	2020-11-02
<u>Country Published:</u>	India

Interested Parties:

Owner:

<u>Person Name</u>	Ankit Sahni
<u>Original Address:</u>	31/42 Punjabi Bagh West New Delhi 110026 India
<u>Current Address:</u>	Same as original address.

Author:

<u>Person Name</u>	RAGHAV Artificial Intelligence Painting App
<u>Original Address:</u>	31/42 Punjabi Bagh West New Delhi 110026 India
<u>Current Address:</u>	Same as original address.

<u>Person Name</u>	Ankit Sahni
<u>Original Address:</u>	31/42 Punjabi Bagh West New Delhi 110026 India
<u>Current Address:</u>	Same as original address.

A RECENT ENTRANCE TO PARADISE



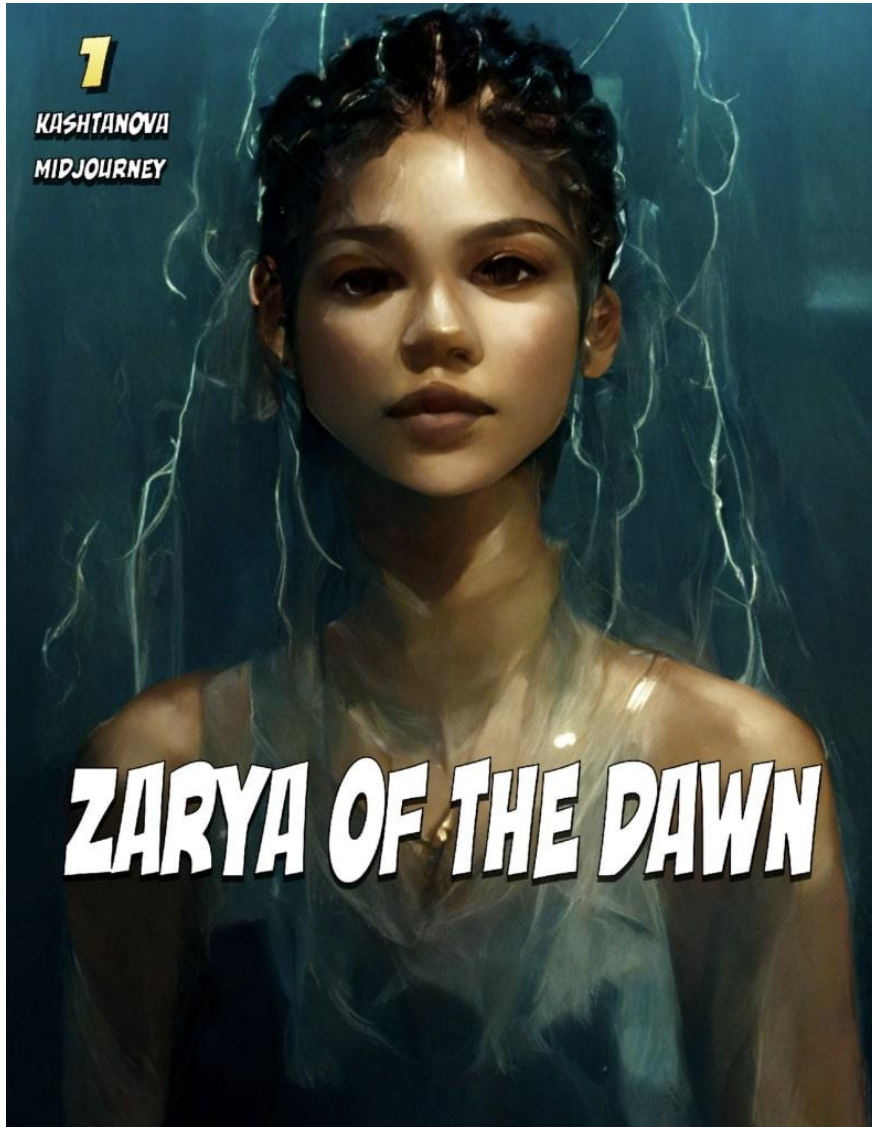
Procedural steps:



Creativity Machine's A Recent Entrance to Paradise

- ***Application for registration***: author: Creativity Machine; owner: Mr. Thaler; concept: work-made-for-hire (November 3, 2018) → **Rejection of registration** (August 12, 2019)
- ***Request of reconsideration*** (September 23, 2019) → **Refused** (March 30, 2020)
- ***Second request of reconsideration*** (May 27, 2020) → **Refused** (February 14, 2022)
 - *Human authorship needed*
 - *Machines have no right to transfer*
 - *WMFH doctrine is on ownership rather than on the existence of copyright protection!*
- ***Case 1:22-cv-01564-BAH*** (August 18, 2023): „defendants are correct that human authorship is an essential part of a valid copyright claim”.

ZARYA OF THE DAWN



Procedural steps:

- **Application for registration:**
author: Kris Kashtanova (+ Midjourney); owner: Ms. Kashtanova → Registration on September 15, 2022
- **Notice of possible cancellation**
on October 28, 2022 → human authorship needed + **disclosure of involvement of external sources / third parties in the creation of the work** → Ms. Kashtanova failed to identify she has created parts of the comic book with Midjourney.

USCO GUIDANCE



UNITED STATES COPYRIGHT OFFICE

Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence

PUBLISHED ON THURSDAY, MARCH 16, 2023

16190 FEDERAL REGISTER, VOL. 88, NO. 51

RULES AND REGULATIONS

37 CFR PART 202

II. The Human Authorship Requirement

In the Office's view, it is well-established that copyright can protect only material that is the product of human creativity. Most fundamentally, the term "author," which is used in both the Constitution and the Copyright Act, excludes non-humans. The Office's registration policies and regulations reflect statutory and judicial guidance on this issue.

THÉÂTRE D'OPÉRA SPATIAL



Procedural steps:

- ***Application for registration:*** author: Jason Allen (no disclosure of the use of Midjourney) → registration denied on December 13, 2022.
- ***Request to reconsider registration:*** USCO misapplied human authorship requirement + public policy favours registration? → denied on June 6, 2023.
- ***Second request to reconsider registration:*** USCO ignored essential human element (series of 600+ prompts) + fair use + value judgment on utility tools → denied on September 5, 2023.
- ***Court proceeding is pending in Colorado*** (Allen v. Perlmutter, Case No. 1:24-cv-02665-WJM).

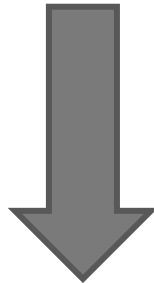
USCO PRACTICE SINCE MARCH 16, 2023



The Register of Copyrights of the United States of America
United States Copyright Office · 101 Independence Avenue SE · Washington, DC 20559-0001

February 23, 2024

*Update of the Register of
Copyright to Chairs and Ranking
Members of Senate and House
committees*



Since the Registration Guidance was issued, the Office's Registration Division has examined hundreds of works that incorporate AI-generated material and has issued registrations to well over 100 so far.¹⁰ Over the past year, the Office's Review Board (the body that issues final agency decisions on registration appeals) has twice upheld refusals to register works created using generative AI.¹¹ In one case, the applicant refused to disclaim

AUTHORSHIP UNDER COPYRIGHT LAW



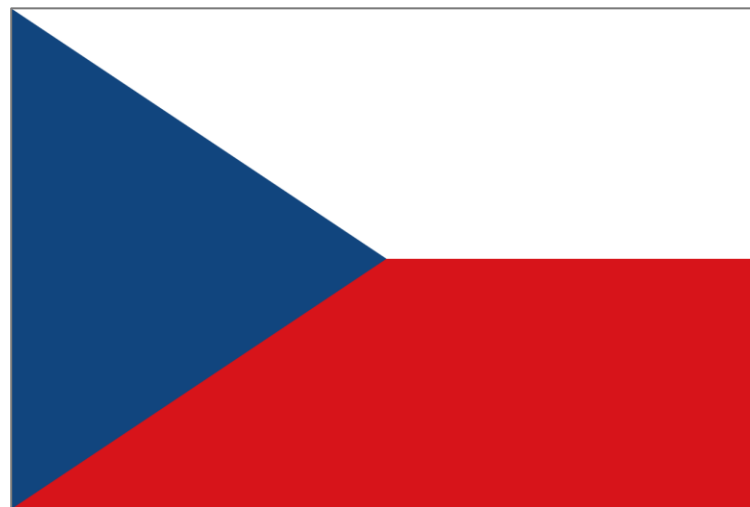
č. j. 10 C 13/2023- 16



ČESKÁ REPUBLIKA

ROZSUDEK

JMÉNEM REPUBLIKY



AUTHORSHIP UNDER COPYRIGHT LAW



AG Munich, February 13, 2026
(142 C 9786/25).



AUTHORSHIP UNDER COPYRIGHT LAW



Nanshan District People's Court in Shenzhen
Tencent v. Shanghai Yinxun
No. (2019) Yue 0305 Min Chu No. 14010

Q: "Does China recognize AI as an author?"

NO.

- Due to human intervention, the **creator of the AI** 'Dreamwriter', i.e., Tencent, **is the author**
- Originality of work
- Granted copyright protection



AUTHORSHIP UNDER COPYRIGHT LAW



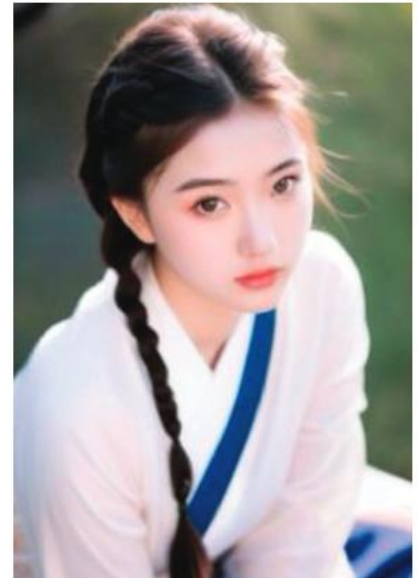
Beijing Internet Court

Lee v. Liu (2023) Jing 0491 Min Chu No. 11279

Q: “Are choices made by individuals in the image generation process creative enough?” + TRANSPARENCY?

“[I]t is the **plaintiff that directly made relevant settings** for the AI model as needed and finally selected the image. That image was generated directly based on the plaintiff’s intellectual input and reflects the plaintiff’s personalized expression. **Therefore, the plaintiff is the author** of the image and enjoys the copyright in that image.

It needs to be clarified that, although this court determines that the plaintiff enjoys the copyright as the author, according to the principle of good faith and the need to protect the public’s right to know, **the plaintiff should clearly mark the AI technology or model he uses**. In this case, the plaintiff marked the image as an ‘AI illustration’, which is sufficient to inform the public that the content was generated by the plaintiff using AI technology. This court affirms that.”



BUT! Plaintiff’s AI-generated images lacked enough original authorship [Feng v. Dongshan Co. (2025) Zhangjiagang Court].



PLATFORM LIABILITY?



Guangzhou Internet Court
Ultraman
(2024) Yue 0192 Min Chu No. 113

Q: “Are platforms liability for the use of their services by their clients (to create outputs that might infringe the copyright of third parties)?

- A generative AI service provider who ***failed to meet a duty of reasonable care*** and whose website allowed for generating pictures identical or similar to the copyrighted image is found liable for copyright infringement of the rights of reproduction, adaptation, and information network dissemination.



AN ARMADA OF PENDING CASES...



DEFENDANTS FACING 1 LAWSUIT

Thomson Reuters v. Ross Intelligence

Dow Jones v. Perplexity

Advance Loc. Media v. Cohere

Andersen v. Stability AI (DeviantArt, Runway, Midj.)

Alcon Entertainment v. Tesla, Elon Musk, Warner Bros.

Huckabee v. Bloomberg

Lehrman v. Lovo

Vacker v. ElevenLabs (DMCA claim)

Pierce v. Photobucket (DMCA claim)

Brave Software v. News Corp. (vol. dismissed)

CanLII v. Caseway AI (Canada)

Kneschke v. LAION (Germany)

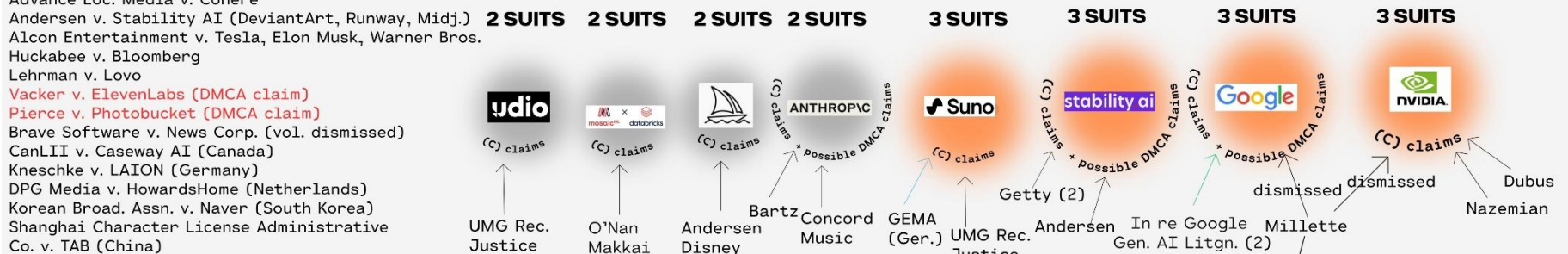
DPG Media v. HowardsHome (Netherlands)

Korean Broad. Assn. v. Naver (South Korea)

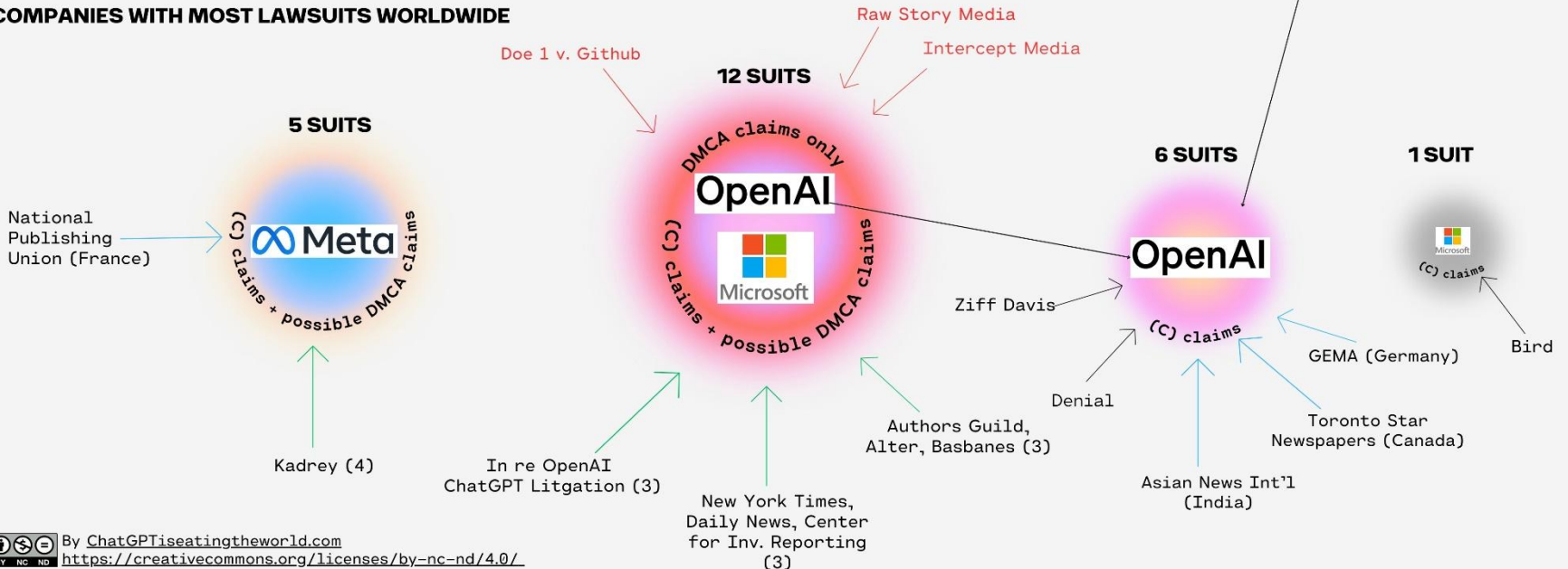
Shanghai Character License Administrative

Co. v. TAB (China)

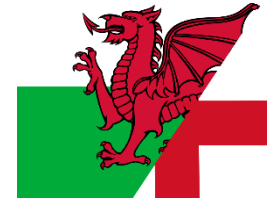
Copyright lawsuits v. AI (worldwide)



AI COMPANIES WITH MOST LAWSUITS WORLDWIDE



AN ARMADA OF (PENDING) CASES...



č. j. 10 C 13/2023- 16

Czech Republic, Germany:
on authorship related
to contents generated
with the help of ChatGPT.



ČESKÁ REPUBLIKA
ROZSUDEK
JMÉNEM REPUBLIKY

UK and Germany: on training GenAI.

Germany, the Netherlands, Hungary, Denmark:
on TDM.

PRESS RELEASE / 21 JANUARY 2025

Fair remuneration demanded: GEMA files lawsuit against Suno Inc.

—

🌐 ✕ in 🔗

GEMA is taking legal action against Suno Inc., an American provider of AI-generated audio content. It is accusing the company of processing protected recordings of world-famous songs from GEMA's repertoire in the tool without paying remuneration. In numerous cases, the AI tool generated audio content that is confusingly similar to the original songs such as 'Forever Young', 'Daddy Cool', 'Mambo No. 5' or 'Cheri Cheri Lady'.

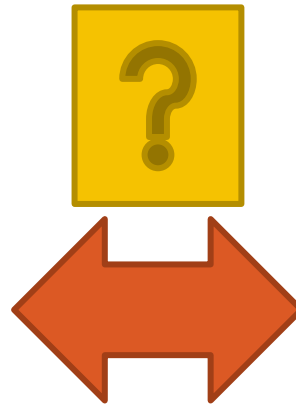
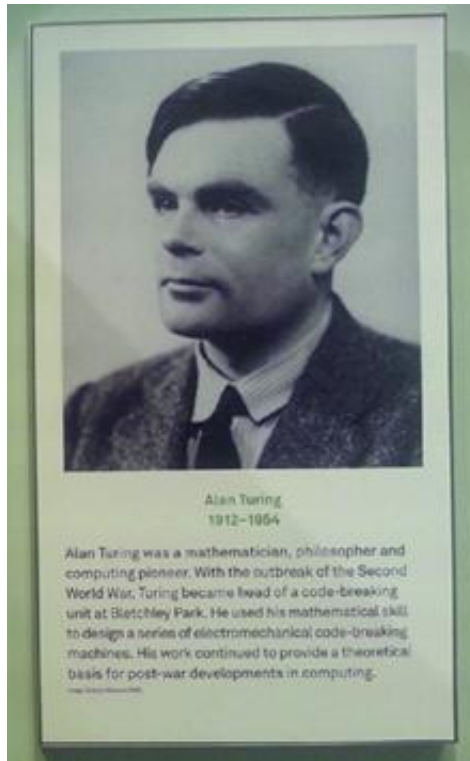
News Report Technology

May 03, 2023

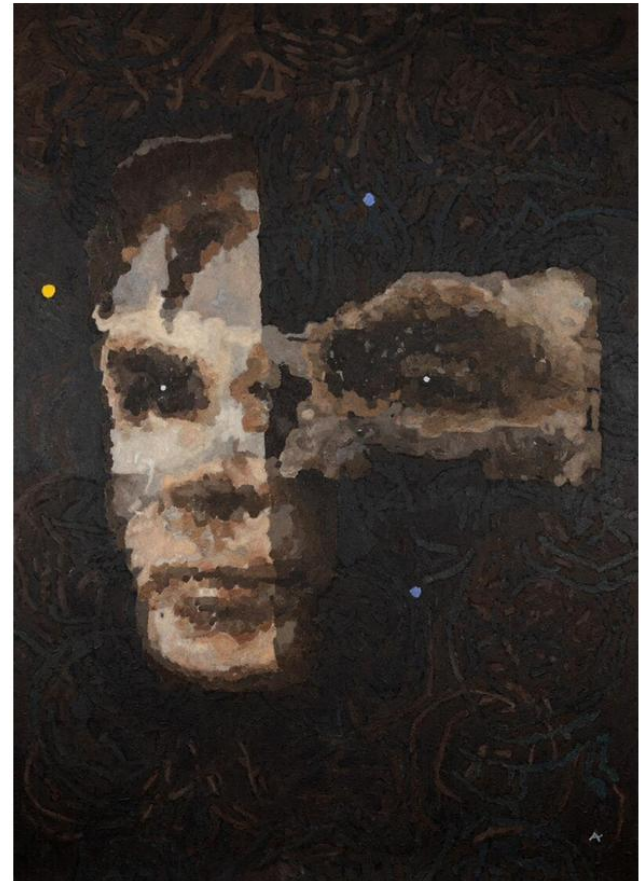
German Stock Photographer Receives Penalty for Request to Remove Photos from Training Dataset

OUTPUT / INFRINGEMENT VIA CONTENT GENERATION?

Reproduction? Adaptation? That is the question...



Ai-DA: A.I.
God –
Portrait of
Alan Turing



OUTPUT / INFRINGEMENT VIA CONTENT GENERATION?

Reproduction? Adaptation? That is the question...

The plaintiffs are wrong to say that, because their books were duplicated in full as part of the LLaMA training process, they do not need to allege any similarity between LLaMA outputs and their books to maintain a claim based on derivative infringement. To prevail on a theory that LLaMA's outputs constitute derivative infringement, the plaintiffs would indeed need to allege and ultimately prove that the outputs "incorporate in some form a portion of" the plaintiffs' books. *Litchfield v. Spielberg*, 736 F.2d 1352, 1357 (9th Cir. 1984); *see also Andersen v.*

Stability AI Ltd., No. 23-CV-00201-WHO, 2023 WL 7132064, at *7-8 (N.D. Cal. Oct. 30, 2023)

Kadrey v. Meta Platforms, order granting motion to dismiss, Case 23-cv-03417-VC (20 November 2023)

OUTPUT / INFRINGEMENT VIA CONTENT GENERATION?

Prima facie infringement? Yes? No?

The Court finds that plaintiffs have plausibly alleged the existence of third-party end-user infringement and that defendants knew or had reason to know of that infringement. Accordingly, plaintiffs have plausibly alleged their contributory copyright infringement claims and defendants' motions to dismiss those claims in all three actions are denied.

The New York Times Co. v. Microsoft Corp., et al., 777 F.Supp.3d 283 (2025), p. 309 and 328.

The alleged abridgments are detailed summaries, usually in bullet point form, of the facts contained in CIR's articles. Those summaries—which differ in style, tone, length, and sentence structure from CIR's articles—are not “substantially similar” to CIR's copyrighted works. They present the “facts in a different arrangement”—bullet point lists or short summary paragraphs—“with a different sentence structure and different phrasing.” *Nihon*, 166 F.3d at 71.

OUTPUT / INFRINGEMENT VIA CONTENT GENERATION?

Prima facie infringement? Yes?

“A more discerning observer could reasonably conclude that the ***allegedly infringing outputs are substantially similar to plaintiffs’ copyrighted works***. Though the ChatGPT-generated summaries submitted by plaintiff do not recount ‘[e]very intricate plot twist and element of character development’ in the original works (...), they are most certainly ***attempts at abridgement or condensation of some of the central copyrightable elements of the original works such as setting, plot, and characters***.

³ OpenAI is mistaken that the outputs at issue in this action are analogous to the summaries of news articles that the Court determined were not substantially similar to copyrighted works in its decision in *N.Y. Times Co. v. Microsoft Corp.*, 777 F. Supp. 3d 283 (S.D.N.Y. 2025). The outputs at issue in that action merely summarized non-copyrightable elements of the original news articles—*i.e.*, “the facts contained in [the news] articles”—and differed “in style, tone, length, and sentence structure” from the original articles. *Id.* at 328. By contrast, the outputs submitted by plaintiffs in this action incorporate such copyrightable elements of plaintiffs’ original works as plot, setting, and characters.

(...) Nothing in this Opinion is intended to suggest a view on whether the allegedly infringing outputs are protected as far uses of the original works.”

OUTPUT / INFRINGEMENT VIA CONTENT GENERATION?

Prima facie infringement? No?

In my judgment, it is not. It is not enough, as it seems to me, that (in Getty Images' words) "the time of making of the copies of the Copyright Works **coincides** with the making of the Model" (**emphasis added**). While it is true that the model weights are altered during training by exposure to Copyright Works, by the end of that process the Model itself does not store any of those Copyright Works; the model weights are not themselves an infringing copy and they do not store an infringing copy. They are purely the product of the patterns and features which they have learnt over time during the training process. Getty Images' central submission that "as soon as it is made, the AI model is an infringing copy" is, accordingly, in my judgment, entirely misconceived. Unlike the RAM chip in *Sony v Ball* which became an infringing copy for a short time, in its final iteration Stable Diffusion does not store or reproduce any Copyright Works and nor has it ever done so. The fact that its development involved the reproduction of Copyright Works (through storing the images locally and in cloud computing resources and then exposing the model weights to those images) is of no relevance. Furthermore, that there is no requirement that an article which is an infringing copy must continue to retain a copy does not assist Getty Images, because it is implicit in the word "continue" that at some point the article has in fact contained an infringing copy. The model weights for each version of Stable Diffusion in their final iteration have never contained or stored an infringing copy.

Getty Images
(US) Inc. et al. v.
Stability AI Ltd.
[2025] EWHC
2863 (Ch)

OUTPUT / INFRINGEMENT VIA CONTENT GENERATION?

Prima facie infringement? **Yes?**

According to the decision of the adjudicating chamber, the plaintiff is entitled to the asserted claims both on the basis of the reproduction of the texts in the language models and their reproduction in the outputs.

Both the memorisation in the language models and the reproduction of the song lyrics in the chatbot's outputs constitute infringements of copyright exploitation rights. These are not covered by any restrictions, in particular the restriction on text and data mining.

Higher Regional Court of Munich – I 42 O 14139/24
Press release dated 11 November 2025 (unofficial translation).
Access via <https://tinyurl.com/4tz6596z>.

INPUT / MACHINE LEARNING – A COPYRIGHT USE?

Machine learning as reproduction

- Szkalej & Senftleben (2024)

3.2 Impact on machine learning

The analysis of ML processes based on the EU position requires a closer look at the individual training and development steps as not every ML stage involves the making of copies. For the sake of simplicity we think of the development phase as involving five stages, consisting of (1) data corpus compilation (2) data corpus preprocessing, (3) data corpus annotation, (4) training of the model, and (5) permanent creation of an artefact (trained model).³²

- Stages 1 and 2 involve reproductions; stages 3 and 4: „the same principles applies: if labelling and training involve copying of copyright-protected data, at least in the operating memory, *these acts are likely to constitute reproductions within the meaning of the broad right of granted in the EU*, even though they are merely a means to an end”. Stage 5 „is not as straightforward to ascertain”.
- And many others: Kraetzig (2024); Hofmann (2024); Moerland (2024); Hamann (2024); Cooper & Grimmelmann (2024); Dornis (2024); Rosati (2024); Quintais (2025); etc.

INPUT / MACHINE LEARNING – A COPYRIGHT USE?

Memorization generally refers to being able to deduce or produce a **model's** given training **example**.

There are further delineations in the literature about different types of memorization. A training **example** may be **memorized** by a model if information about that training example can be **discovered** inside the model through any means. A training example is said to be **extracted** from a model if that model can be prompted to generate an output that looks exactly or almost exactly the same as the training example. A training example may be **regurgitated** by the model if the generation looks very similar or almost exactly the same as the training example (with or without the user's intention to extract that training example from the model).

To tease these words apart: a training example is **memorized** by a model and can be **regurgitated** in the generation process regardless of whether the intent is to **extract** the example or not.

The word memorization itself may be used to refer to other concepts that we may colloquially understand as “memorization.” For example, facts and style (artists style) may also be memorized, regurgitated, and extracted. However, this use should not be confused with technical words (e.g., **extraction**) with precise definitions that correspond to metrics.

OpenAI:
“*Regurgitation*”
is a rare bug
that we are
working to
drive to zero.

Source:
[https://openai.com/
blog/openai-and-
journalism#OpenAI](https://openai.com/blog/openai-and-journalism#OpenAI)

<https://genlaw.org/glossary.html#memorization>

INPUT / MACHINE LEARNING – A COPYRIGHT USE?

Various algorithms memorize various training data

- Ziegler (2021) & Ippolito et al. (2022) on computer codes; Somepalli et al. (2022) & Carlini et al. (2023) on text-to-image models; Chang et al. (2023) & Duarte et al. (2024) on literary works; Karamolegkou et al. (2023) on literary works and LeetCode problem descriptions [+ Wei et al. (2024) on security & privacy] = **memorization exists, but its rate is low and its rate varies in categories of source works.**

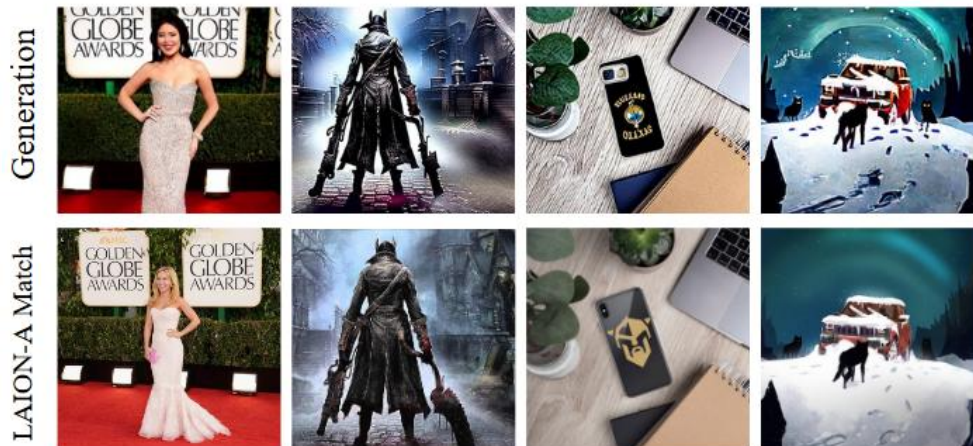


Figure 1. *Stable Diffusion* is capable of reproducing training data, creating images by piecing

Somepalli et al. (2022)

Source	GPT-4	ChatGPT
pre-1923 LitBank	0.244	0.072
Genre: SF/Fantasy	0.235	0.108
Genre: Horror	0.054	0.028
Bestsellers	0.033	0.016
Genre: Action/Spy	0.032	0.007
Genre: Mystery/Crime	0.029	0.014
Genre: Romance	0.029	0.011
Pulitzer	0.026	0.011
Global	0.020	0.009
BBIP/BCALA	0.017	0.011

Table 2: Name cloze performance by book category.

Chang et al. (2023)

INPUT / MACHINE LEARNING – A COPYRIGHT USE?

Various algorithms memorize various training data

- Sag (2023) == “memorization is more likely if a text description is closely associated with a particular image over and over again; it is also more likely if the image is relatively simple or relates to a single subject”



INPUT / MACHINE LEARNING – A COPYRIGHT USE?

Uncertainties around machine learning – to copy or not to copy? That is the question...

14 2. Stability downloaded or otherwise acquired copies of billions of copyrighted
15 images without permission to create Stable Diffusion, including Plaintiffs'. These images are
16 defined below as "Training Images."

17 3. By training
18 to be stored at and incorporated
19 without the consent of the

20 4. When used
21 Training Images to produce
22 These "new" images are
23 particular images Stable I
24 merely a complex collage

13 12. Copilot is run entirely on Microsoft's Azure cloud-computing platform.
14 13. Copilot often simply reproduces code that can be traced back to open source
15 repositories or open-source licensees. Contrary to and in violation of the Licenses, code
16 reproduced by Copilot *never* includes attributions to the underlying authors.

17 14. GitHub and OpenAI have offered shifting accounts of the source and amount of
18 the code or other data used to train and operate Copilot. They have also offered shifting
19 justifications for why a commercial AI product like Copilot should be exempt from these license
20 requirements, often citing "fair use."

J. Doe 1 et al. v. GitHub Inc., et al., plaintiff's
complaint, Case No. 22-cv-06823-JST
(3 November, 2022)

Andersen et al. v. Stability AI et al., plaintiff's
complaint, Case 3:23-cv-00201 (13 January, 2023)

INPUT / MACHINE LEARNING – A COPYRIGHT USE?

To copy or not to copy? That is the question...

64. This commercial success is built in large part on OpenAI's large-scale copyright infringement. One of the central features driving the use and sales of ChatGPT and its associated products is the LLM's ability to produce natural language text in a variety of styles. To achieve this result, OpenAI made numerous reproductions of copyrighted works owned by The Times in the course of "training" the LLM.

Exhibit J

The New York Times Co. v. Microsoft Corp., OpenAI, plaintiff's complaint, 1:23-cv-11195 (Dec. 27, 2023)

The New York Times Co. v. Microsoft Corp., et al., 777 F.Supp.3d 283 (2025), p. 297.

X. CONCLUSION

For the foregoing reasons, the Court denies (1) OpenAI's motions to dismiss the direct infringement claims involving conduct occurring more than three years before the complaints were filed; (2) defendants' motions to dismiss the contributory copyright infringement claims; and (3) defendants' motions to dismiss the state and federal trademark dilution claims in the *Daily News* action.

UMG Recordings, Inc., et al. v. Suno Inc., et al., plaintiff's complaint, 1:24-cv-11611 (June 24, 2024). Identical to this, see UMG Recordings, Inc., et al. v. Uncharted Labs, Inc., et al., plaintiff's complaint, 1:24-cv-11611 (June 24, 2024) §52 et seq.

INPUT / MACHINE LEARNING – A COPYRIGHT USE?

To copy or not to copy? That is the question...

42 (a) Der Bekl. hat die Vervielfältigungshandlung zum Zwecke der Gewinnung von Informationen über „Korrelationen“ im Wortsinn des § 44b I UrhG vorgenommen. Der Bekl. hat das streitgegenständliche Lichtbild von seinem ursprünglichen Speicherort heruntergeladen, um mittels einer bereits verfügbaren Software den Bildinhalt mit der zu dem Text bereits hinterlegten Bildbeschreibung abzugleichen. Diese Analyse der Bilddatei zum Abgleich mit einer vorbestehenden Bildbeschreibung stellt ohne weiteres eine Analyse zum Zwecke der Gewinnung von Informationen über „Korrelationen“ (nämlich der Frage der Nicht-/Übereinstimmung von Bildern und Bildbeschreibungen) dar. Dass der Bekl. die in den Datensatz aufgenommenen Bilder auf diese Art und Weise analysiert hat, wurde klägerseits als solches nicht bestritten.

INPUT / MACHINE LEARNING – A COPYRIGHT USE?

To copy or not to copy? That is the question...

GEMA is taking legal action against Suno Inc., an American provider of AI-generated audio content. It is accusing the company of processing protected recordings of world-famous songs from GEMA's repertoire in the tool without paying remuneration. In numerous cases, the AI tool generated audio content that is confusingly similar to the original songs such as 'Forever Young', 'Daddy Cool', 'Mambo No. 5' or 'Cheri Cheri Lady'.

Source:
[GEMA](#)

To enforce the rights of its members, GEMA filed a lawsuit against the AI provider Suno Inc. with the Munich Regional Court on 21 January 2025. The music tool makes it possible to generate playable audio content using simple prompts. **GEMA was able to document that the system outputs content that obviously infringes copyrights.** In terms of melody, harmony and rhythm, this content largely corresponds to world-famous works whose authors GEMA represents. The songs affected include those by Alphaville (Forever Young), Kristina Bach (Atemlos), Lou Bega (Mambo No. 5), Frank Farian (Daddy Cool) and Modern Talking (Cheri Cheri Lady).

According to the decision of the adjudicating chamber, the plaintiff is entitled to the asserted claims both on the basis of the reproduction of the texts in the language models and their reproduction in the outputs.

Both the memorisation in the language models and the reproduction of the song lyrics in the chatbot's outputs constitute infringements of copyright exploitation rights. These are not covered by any restrictions, in particular the restriction on text and data mining.

Higher Regional Court of Munich, Press release dated. Access via <https://tinyurl.com/4tz6596z>.

INPUT / MACHINE LEARNING – A COPYRIGHT USE?

To copy or not to copy? That is the question...

- 7 An article appeared on one of the applicant's protected online press publications (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.
- 8 In response to the question 'Can you provide a summary in Hungarian of the online press publication that appeared on 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.

INPUT / MACHINE LEARNING – ANY EXCEPTION FOR USERS?

Assume the presence of use – any exception at hand?

Jurisdiction matters (again)

- US: **fair use** (Google Books/scraping case law v. post-Warhol?)
- EU: closed list of E&Ls:
 - **permissionless observing, study and testing** [§5(3) Software D],
 - **temporary acts of reproduction** [§5(1) InfoSoc] &
 - **commercial TDM** [§4 CDSMD],
 - **pastiche** [§5(3)(k) InfoSoc + §17(7) CDSMD]?

INPUT / MACHINE LEARNING – ANY EXCEPTION FOR USERS?



Thomson Reuters v. Ross Intelligence, 765 F.Supp.3d 382 (2025): the purposeful reproduction of copyright protected headnotes from Westlaw is *unfair use*.

Bartz v. Anthropic (2025) model training and building a library from publicly available copyright protected materials is „**exceedingly transformative**”; but (1) infringing outputs generated by users could lead to a different conclusion; and (2) building a training data set from shadow libraries is *unfair use* → 1.5 billion USD settlement!

Kadrey v. Meta (2025): training Llama was „**highly transformative**” and plaintiffs failed to prove negative market harms. The court failed to address the use of shadow libraries by Meta.

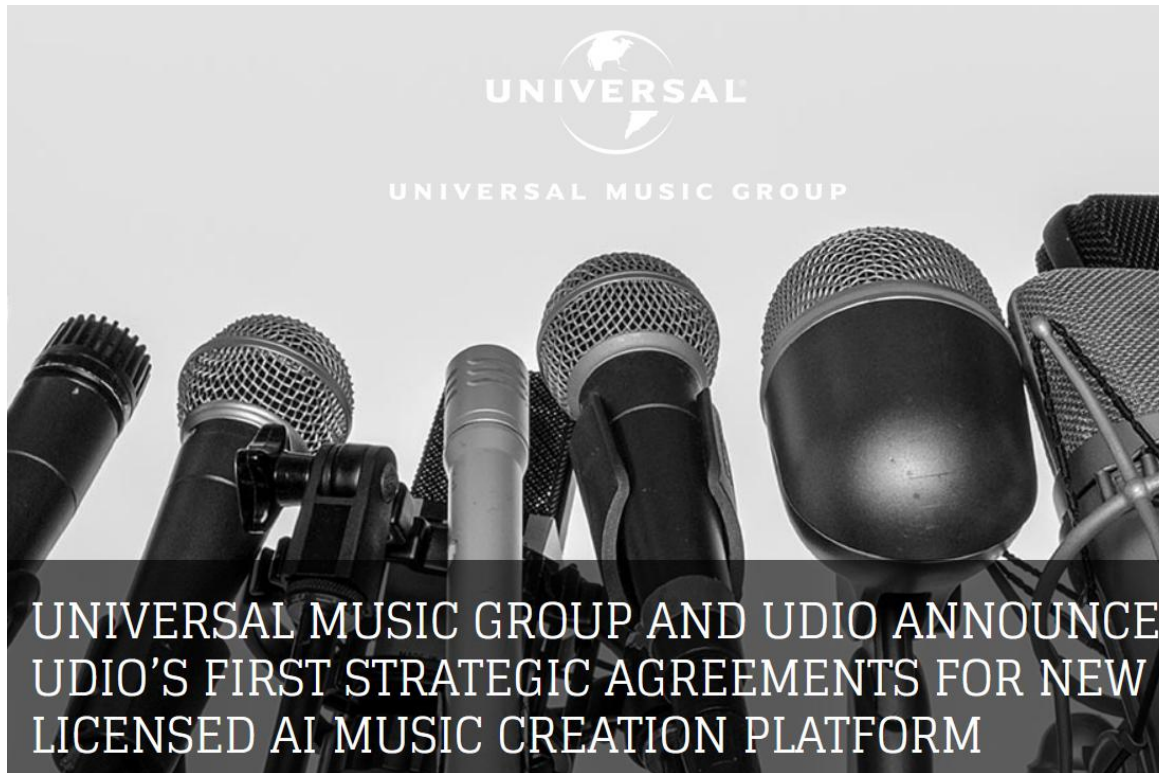
INPUT / MACHINE LEARNING – ANY EXCEPTION FOR USERS?

If approved, this landmark settlement will be the largest publicly reported copyright recovery in history, larger than any other copyright class action settlement or any individual copyright case litigated to final judgment. The proposed settlement well surpasses other copyright recoveries, will provide meaningful compensation for each Class Work, and will set a precedent of AI companies paying for their use of pirated websites like Library Genesis and Pirate Library Mirror. Subject to court approval, the principal terms are:

- Anthropic will pay the Class at least ***\$1.5 billion dollars***, plus interest. With around 500,000 works in the Class, this amounts to an estimated gross recovery of ***\$3,000 per Class Work***.
- Anthropic will ***destroy the LibGen and PiLiMi datasets*** after the expiration of any litigation preservation or other court orders.
- In exchange, Anthropic will receive a ***past release only*** for conduct up to August 25, 2025. Claims arising out of conduct after August 25, 2025 won't be released by the Settlement, nor will any claims (past or future) arising out of allegedly infringing outputs from Anthropic's AI models.

This result is nothing short of remarkable.

INPUT / MACHINE LEARNING – ANY EXCEPTION FOR USERS?



See:
<https://www.universalmusic.com/universal-music-group-and-udio-announce-udios-first-strategic-agreements-for-new-licensed-ai-music-creation-platform/>
(October 29, 2025)

In addition to the compensatory legal settlement, the new license agreements for recorded music and publishing will provide further revenue opportunities for UMG artists and songwriters.

The new platform, which will be launched in 2026, will be powered by new cutting-edge generative AI technology that will be trained on authorized and licensed music. The new subscription service will transform the user engagement experience, creating a licensed and protected environment to customize, stream and share music responsibly, on the Udio platform.

INPUT / MACHINE LEARNING – ANY EXCEPTION FOR USERS?

The Walt Disney Company and OpenAI Reach Landmark Agreement to Bring Beloved Characters from Across Disney's Brands to Sora

BRAND | DECEMBER 11, 2025 | 6 MIN. READ

- As part of this three-year licensing agreement, Sora will be able to generate short, user-prompted social videos that can be viewed and shared by fans, drawing on more than 200 Disney, Marvel, Pixar and Star Wars characters.
- Agreement will make a selection of these fan-inspired Sora short form videos available to stream on Disney+.
- Disney and OpenAI affirm a shared commitment to responsible use of AI that protects the safety of users and the rights of creators.
- Alongside the licensing agreement, Disney will become a major customer of OpenAI, using its APIs to build new products, tools, and experiences, including for Disney+, and deploying ChatGPT for its employees.
- As part of the agreement, Disney will make a \$1 billion equity investment in OpenAI, and receive warrants to purchase additional equity.

See:

<https://thewaltdisneycompany.com/news/disney-openai-sora-agreement/>

(December 27, 2025)

INPUT / MACHINE LEARNING – ANY EXCEPTION FOR USERS?

§4 CDSMD – shady contours / open questions

- Is machine learning “generation of information such as patterns, trends, correlations”?
- Late implementations + divergencies in the statutory language in the MSs, e.g.
 - exception or limitation;
 - “lawfully accessible” works;
 - retaining copies “as long as necessary”;
 - reservation of rights via “computer-readable means”;
 - more on opt-out...



INPUT / MACHINE LEARNING – ANY EXCEPTION FOR USERS?

§4(3) CDSMD – shady contours / open questions

- Ex ante (source page-level) or ex post (training data-level) reservation of rights? (Compare to recital 18 CDSMD.)
General or specific? Reservation via platforms or directly by rightsholders?
 - Can ex ante or ex post opt-outs be effective at all?
 - What about repeated posts of contents online?
 - If the single removals are ineffective, will AI service providers face a “stay-down” obligation to moderate their datasets?
- What is computer readable form? What contractual agreements? What unilateral declarations?
 - Can a TPM or an RMI exclude the applicability of rights reservation?

FIRST REQUEST FOR A PRELIMINARY RULING



EU/Hungary

Summary C-250/25 – 1

Case C-250/25

Summary of the request for a preliminary ruling pursuant to Article 98(1) of the Rules of Procedure of the Court of Justice

Date lodged:

3 April 2025

Referring court:

Budapest Környéki Törvényszék (Hungary)

Date of the decision to refer:

10 March 2025

Applicant:

Like Company

Defendant:

Google Ireland Limited

Questions referred for a preliminary ruling

1. Must Article 15(1) of Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 [on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC], and Article 3(2) of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, be interpreted as meaning that the display, in the responses of an LLM-based chatbot, of a text partially identical to the content of web pages of press publishers, where the length of that text is such that it is already protected under Article 15 of Directive 2019/790, constitutes an instance of communication to the public? If the answer to that question is in the affirmative, does the fact that [the responses in question are] the result of a process in which the chatbot merely predicts the next word on the basis of observed patterns have any relevance?
2. 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?
3. 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?
4. Must Article 15(1) of Directive 2019/790 and Article 2 of Directive 2001/29 be interpreted as meaning that, where a user gives an LLM-based chatbot an instruction which matches the text contained in a press publication, or which refers to that text, and the chatbot then generates its response based on the instruction given by the user, the fact that, in that response, part or all of the content of a press publication is displayed constitutes an instance of reproduction on the part of the chatbot service provider?

AI ACT – AN EXTRA LAYER

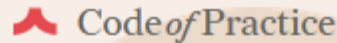
Article 53

Obligations for providers of general-purpose AI models

1. Providers of general-purpose AI models shall:
 - (c) put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790;
 - (d) draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office.

+ **Transparency requirements** per §53(1)(a)-(b), except for FOSS based general-purpose AI models (see further recital 104 & 106);
+ **Codes of practice** until a standard is published per §53(4).

AI ACT – AN EXTRA LAYER



SUMMARY TRANSPARENCY COPYRIGHT SAFETY & SECURITY

🔗 Copyright Section

🔗 Commitment I.2. Copyright policy

In order to fulfil the obligation to put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to [Article 4\(3\) Directive \(EU\) 2019/790](#) pursuant to Article 53(1), point (c) AI Act, Signatories commit to drawing up, keeping up-to-date, and implementing a copyright policy in accordance with Measure [I.2.1](#), as well as adopting Measures [I.2.2–I.2.6](#) for their general-purpose AI models placed on the EU market.

[Read commitment and measures](#)

🔗 II. Commitments by Providers of General-Purpose AI Models with Systemic Risk

AI ACT – AN EXTRA LAYER



Brussels, 24.7.2025
C(2025) 5235 final

ANNEX

(15) The Template annexed to this Explanatory Notice aims to provide a common minimal baseline for the information to be made publicly available in the Summary. It consists of three main sections:

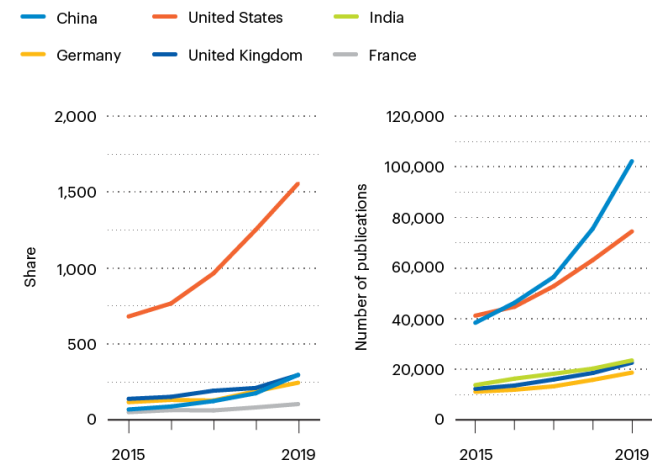
1. **General information:** this section requires information allowing identification of the provider and of the model, and information on modalities, the size of each modality within broad ranges, as well as general characteristics of the training data.
2. **List of data sources:** this section requires disclosure of the main datasets that were used to train the model, such as large private or public databases, and a comprehensive narrative description of the data scraped online by or on behalf of the provider (including a summary of the most relevant domain names scraped) and a narrative description of all other data sources used (e.g. user data or synthetic data) to ensure completeness of the summary regarding the content used for the model training¹⁵.
3. **Relevant data processing aspects:** this section of the Template requires disclosure of certain data processing aspects that are relevant for the exercise of the rights of parties with legitimate interests under Union law. This is especially important for compliance with Union law on copyright and related rights and for the removal of illegal content to mitigate the risk that such illegal content may be reproduced and disseminated at scale by the general-purpose AI model.

CONCLUSION: ON THE PROTECTABILITY OF AI-GENERATED OUTPUTS

- ***Lack or existence of formalities***
 - copyright v. patent law (different purposes of law AND of registration)
 - protection $\neq$ registration (but overlapping concepts)
 - formalities = formal and substantive requirements
 - concurring/conflicting national approaches
- ***Human or deemed authorship / inventorship?***
 - Policy rather than philosophy!
 - Is it really the machine that requests protection? Who are the key beneficiaries?
 - Transfer of rights in the lack of e-personality?
- ***Global race for AI protection!***

STEEP COMPETITION

The five leading locations for AI-related output in the Nature Index (left) and Dimensions (right). In both databases, China achieved the fastest growth in output, followed by the UK in Dimensions, and the US in the Nature Index. India is the 3rd most productive location for AI research across all fields in the Dimensions database, but appears at 20th position for AI-related natural-sciences research in the Nature Index.





THANKS FOR YOUR ATTENTION!

mezei.peter@szte.hu
<http://ssrn.com/author=1697918>