

The risks of AI slop and AI model collapse, and why we need additional incentives and a dual right system to adequately feed the next Gen AI models and to remunerate creators

Alain Strowel

Professor, UCLouvain, Magister in IP/IT Alicante, Munich IP Law Center

Attorney, Pierstone, Brussels

alain.strowel@uclouvain.be / alain.strowel@pierstone.com

- Co-author with Luc Desaunettes and Sari Depreeuw of a study on AI tools and their impact on the creative sector: *Study on copyright and new technologies: copyright data management and artificial intelligence*, Publications Office of the European Union, March 2022, <https://data.europa.eu/doi/10.2759/570559> / <http://hdl.handle.net.proxy.bib.ucl.ac.be/2078/266347>



The Great Backlash (demonstrations in Hollywood)

The Guardian
(25/2/2025)

Why are creatives fighting UK government AI proposals on copyright?

More than 48,000 creators including Kate Bush and Abba's Björn Ulvaeus say Labour's preferred option is unfair and unworkable



📷 The latest protest against the government AI proposals is a silent album co-written by more than 1,000 musicians, including Kate Bush. Photograph: John Carder Bush

Outline

- Shifts in value creation and substitutive effects
- The risks (if no corrective measure):
 - Continuing legal uncertainties
 - AI slop
 - AI model collapse
- A bold defense of a dual system with reinforced incentives for human-generated data
 - Improved market solution
 - New unwaivable right of remuneration for creators

Uncertain shifts in jobs and value creation...

- Study of the trends in demand for online freelancers (based on 1.388.711 (!) job posts from July 2021 to July 2023)
 - *“the introduction of ChatGPT and image-generating tools led to nearly immediate decreases in posts for online gig workers across job types, but particularly for automation-prone jobs”*
- But: up- and re-skilling of workers
 - the integration of AI tools “into work is becoming increasingly valued”
- GenAI has the power to displace some creative workers. So what?
 - Should we care? Is tech innovation not always, like capitalism, about « creative destruction » (Schumpeter)?
 - Should we care more than for the displacement of other jobs (white collar jobs...)?
 - Shouldn't we be more optimistic about the future integration of those tools with new creative practices? Can we expect a new creative empowerment by AI?
 - Comparison with the threat of photography to (portrait) painting in the late 19th century?

...but clear substitutive effect

“A large language model trained on short stories will be able to create competing short stories. An AI image model trained on stock images will be able to create competing stock images. An AI music model trained on music that’s licensed to TV shows will be able to create competing music to license to TV shows. And these models, even if they’re imperfect, can be used so quickly and cheaply that this competition is inevitable” (Ed Newton-Rex)

- Reduction in demand (and readiness to pay) for human content
 - Displacement for (some) creative people who are not in a position to up- and re-skill rapidly
 - Attention must be focused on the creators who have been used to write short stories, to make stock images, to compose music for TV shows, etc.

Focus of the law on those who will not easily transition and are more vulnerable

Need for corrective measures (question of redistributive justice)

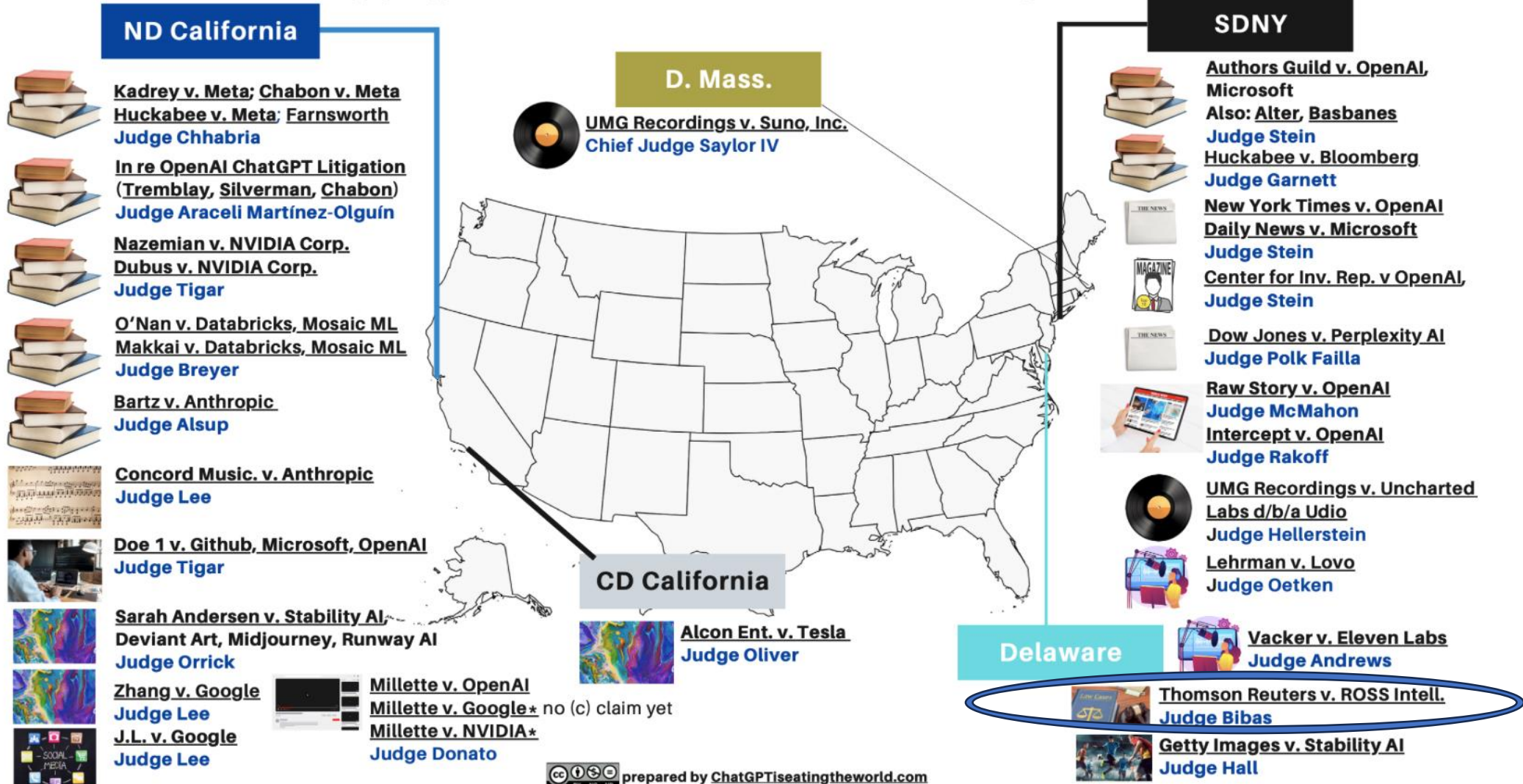
- Some form of remuneration for those mostly affected

Need for stronger incentives to create human works and data to put in place by legislation

- Because of risks:
 - Continuing legal uncertainties and divergences
 - AI slop
 - AI model collapse

US: Risk of increasing divergences and uncertainties

Copyright + Related Lawsuits v. AI Companies



Various claims with uncertain outcome (and moral dimension)

- Copyright claims: derivative works + DMCA + fair use
 - *Is fair use not simply the right to hire a lawyer and take one's chances in court?*
- Unfair competition (California Act)
- Unjust enrichment → idea of (unjust) taking
 - Actual damages from the deprivation of the benefits of the works v. derived profit and economic benefits = inequitable and unjust

Thomson Reuters v. Ross Intelligence

(summary judgment, 11 Febr. 2025)

Feist Publications, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 345 (1991).

Question

Does originality for copyright purposes mean that the work was independently created and has some minimal degree of creativity?

West Headnote

Originality, for copyright purposes, means that the work was independently created and has some minimal degree of creativity.

Case Opinion

Original, as the term is used in copyright, means only that the work was independently created by the author (as opposed to copied from other works), and that it possesses at least some minimal degree of creativity.

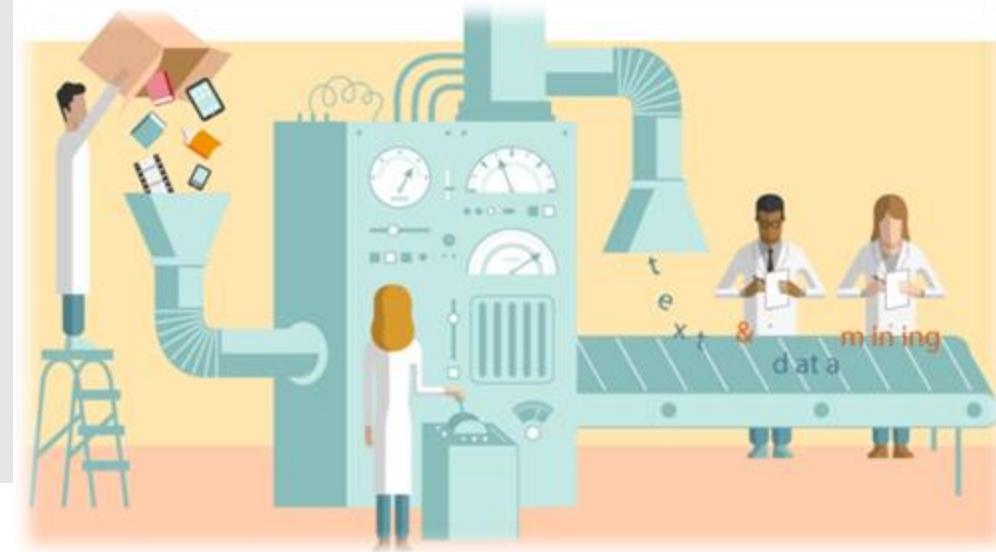
Thomson Reuters v. Ross Intelligence

- Defense of fair use with burden of proof for Ross/defendant and 4 factors:
 - Purpose and character of use FOR THOMSON: commercial, not transformative (« because « it does not have a « further purpose of different character » from Thomson Reuters's »)
 - Intermediate copies : allowed for computer program (Google v. Oracle, 593 U.S.= SCOTUS) but then « copying was necessary for competitors to innovate » or for reverse engineer/interoperability
 - Nature of original work FOR ROSS: not that creative
 - How much was used/how substantial FOR ROSS: what matters = « not « the amount and substantiality of the portion used in making the copy, but rather the amount and substantiality of what is thereby made accessible to a public » → based on the similarity in outputs: « competing substitute » or not »
 - Effect on the market/value of work (most important factor) FOR THOMSON
 - Including on the potential derivative markets = « **data to train legal AIs** »
 - « Public interest in accessing the law. But legal opinions are freely available

EU: Exceptions for text and data mining (TDM)

Art. 2(2) 2019/790 CDSM Dir.

Text and data mining" is defined as "any automated analytical technique for analyzing text and data in digital form to **generate information** that includes, but is not limited to, **patterns, trends and correlations**"



Small doubts as to the application to GenAI:

<https://www.copyrightuser.org/understand/exceptions/text-data-mining/>

- Textual argument: what about the output of GenAI? is it information? Or rather content or work (would be © if author)
- Argument based on *Travaux préparatoires*: no reference to GenAI
- If it applies to GenAI, to what exactly? Whole AI training or AI life cycle?

Legal uncertainties in the 2 TDM exceptions

For research (art. 3)

- For research organizations & cultural heritage institutions
 - Not for profit
- Condition: **legal access to the works**
- Mandatory exception >< cannot be set aside by contract (s. 7(1))
- **No opt-out**

For other beneficiaries (art. 4)

- For everyone
- For any purpose
 - including commercial
- Condition: **legal access to the works**
- Exception mandatory but may be waived by contract (s. 7(1))
- **Except if opt-out (reservation)**: right holders can reserve the use "**in an appropriate manner**" = what about it?
 - robot.txt or other machine-readable tag (not enough T&C...)

Room for a remuneration in case of TDM?

- **For the Art. 3 TDM:** Recital 17 of CDSMD excludes it because « any potential harm created to rightholders through this exception would be minimal »
 - « Member States should, therefore, not provide for compensation for rightholders » in this case
- **For the Art. 4 TDM:** no reference to the de minimis harm threshold
→ opening for remuneration system?
- But this framework concerns the input/training phase and nothing about the output/deployment phase

Contradicts the reading of Art. 3 by ChatGPT: amazing!

With respect to the use of copyrighted data to train AI systems, the European Union recently adopted the Directive on Copyright in the Digital Single Market, which contains an **exception** for the use of certain copyrighted works **for text and data mining** purposes, which is used to train AI systems.

However, this exception is limited to scientific research purposes and is subject to certain conditions, such as the obligation to provide equitable remuneration to the relevant right holders.



peux-tu comparer l'application des exceptions au droit d'auteur en droit américain et en droit européen au cas où un outil d'intelligence artificielle utilise des données de départ pour entraîner le système, parmi lesquelles on trouve des données comme des photos ou du texte qui sont protégés par un droit d'auteur?



Aux États-Unis, le droit d'auteur contient une exception de "fair use" qui permet l'utilisation d'œuvres protégées par le droit d'auteur à des fins telles que la critique, les commentaires, l'enseignement, les rapports d'actualités ou la recherche scientifique. Cette exception est appliquée en cas d'utilisation d'œuvres protégées par le droit d'auteur dans un contexte d'IA pour entraîner un système d'IA. La Cour suprême américaine a également récemment confirmé dans l'affaire Google c. Oracle que l'utilisation de certaines parties d'une œuvre protégée par le droit d'auteur à des fins d'interopérabilité avec d'autres logiciels était un "fair use".

En Europe, le droit d'auteur contient également certaines exceptions telles que la citation, la parodie, la caricature et l'enseignement, mais ces exceptions sont plus limitées que la notion de "fair use" aux États-Unis. En ce qui concerne l'utilisation de données protégées par le droit d'auteur pour entraîner des systèmes d'IA, l'Union européenne a récemment adopté la Directive sur le droit d'auteur dans le marché unique numérique, qui contient une exception pour l'utilisation de certaines œuvres protégées par le droit d'auteur à des fins de text et data mining, qui est utilisé pour entraîner des systèmes d'IA. Cependant, cette exception est limitée aux fins de recherche scientifique et est soumise à certaines conditions, telles que l'obligation de fournir une rémunération équitable aux titulaires de droits concernés.

Plainly wrong. Not sure that OpenAI is of the same opinion!

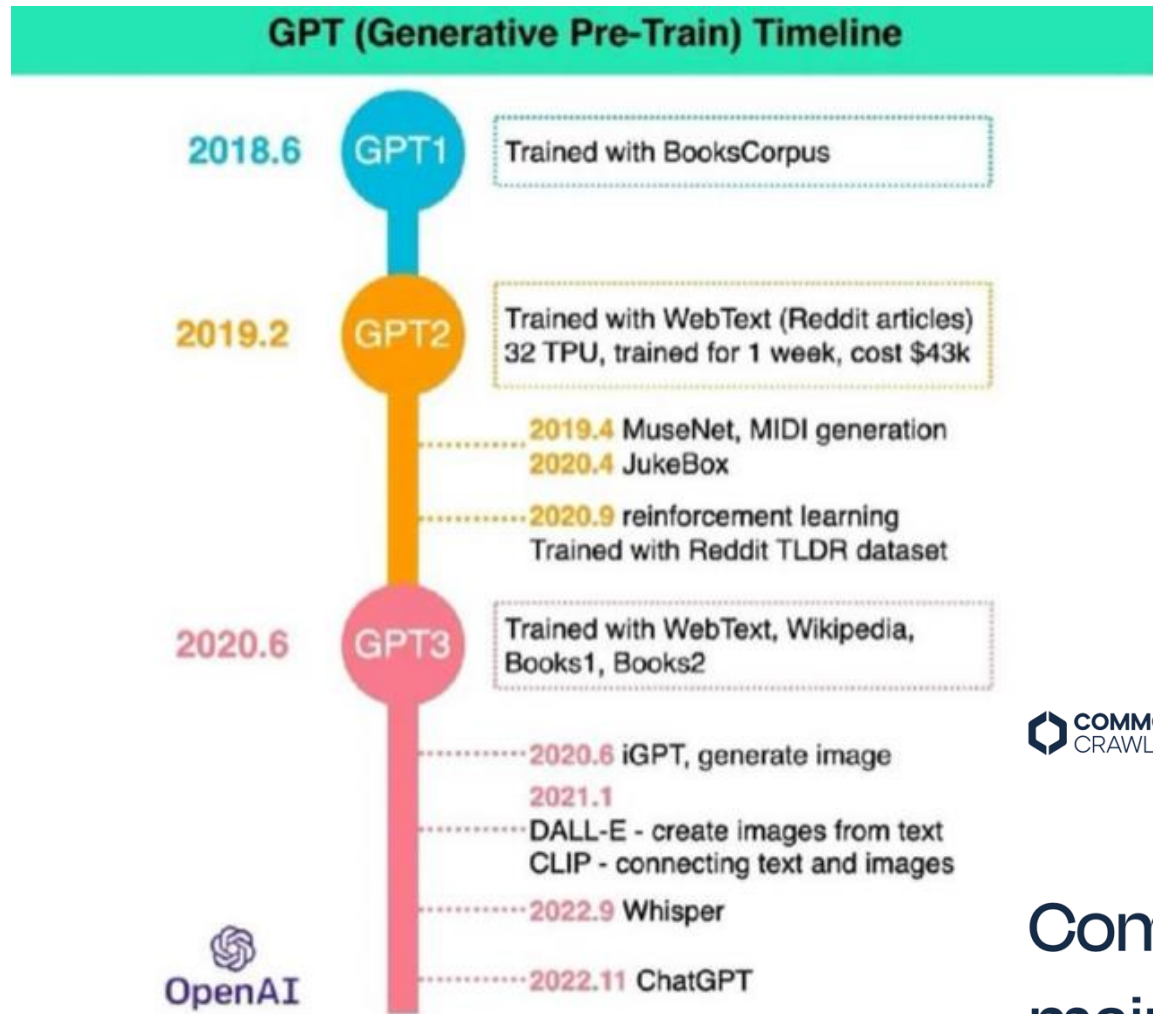
Timeline for final decision in copyright wars? For a solution to the opt-out conundrum?

- U.S.:
 - If like in the *Authors Guild v. Google Books* fair use case: still 8-10 years?
 - Action filed in 2005, first instance decision in 2013, appeal in 2015 and Supreme Court refusal to consider the case in 2016
- EU:
 - no expectation of rapid clarification as to the application of the TDM exception to GenAI and to the (whole) process of AI training / development/life cycle
 - no foreseeable solution on the standardisation of the opt-outs
 - in the meantime, one can anticipate diverging national solutions
 - Example of Spain with an Extended Collective License model (whitelisted by Art. 12 CDSMD)
- **Not satisfactory!**

Risk of 'AI slop'

- 2024 was the year of “**AI slop**” (a term that has acquired its “lettre de noblesse” as it now appears as entry on Wikipedia)
 - AI slop might be defined as “low-quality media” produced by GenAI tools
 - It refers to “vague text, filled with buzzwords, with no real point”
 - = « AI Garbage », nears the term « spam »
 - But it attracts higher ad revenues
 - New feature announced by Meta: AI characters will take over Instagram and Facebook – and AI slop will get on steroids
- Implications for AI training that is largely done from online sources

Black boxes:
what training
data?



Ex. of sources: Wikipedia, Reddit,
Common Crawl



The Data ▾ Resources ▾

Common Crawl
maintains a **free, open
repository** of web crawl
data that can be used by
anyone.

Risk of AI model collapse

- **“AI model collapse”**: when AI systems trained primarily on synthetic data start losing the diversity and quality of human-generated data
 - Using synthetic data permits to bypass some of the issues with real data: data and copyright protections
 - Amplification of the initial biases and inaccuracies can lead to the propagation of misalignments and the degradation of the model’s performance
- Human-linked data thus seems to increase in value for AI training in so far there is a growing use and online dissemination of synthetic or AI-generated data
- At this junction the copyright-based incentive system might play a role

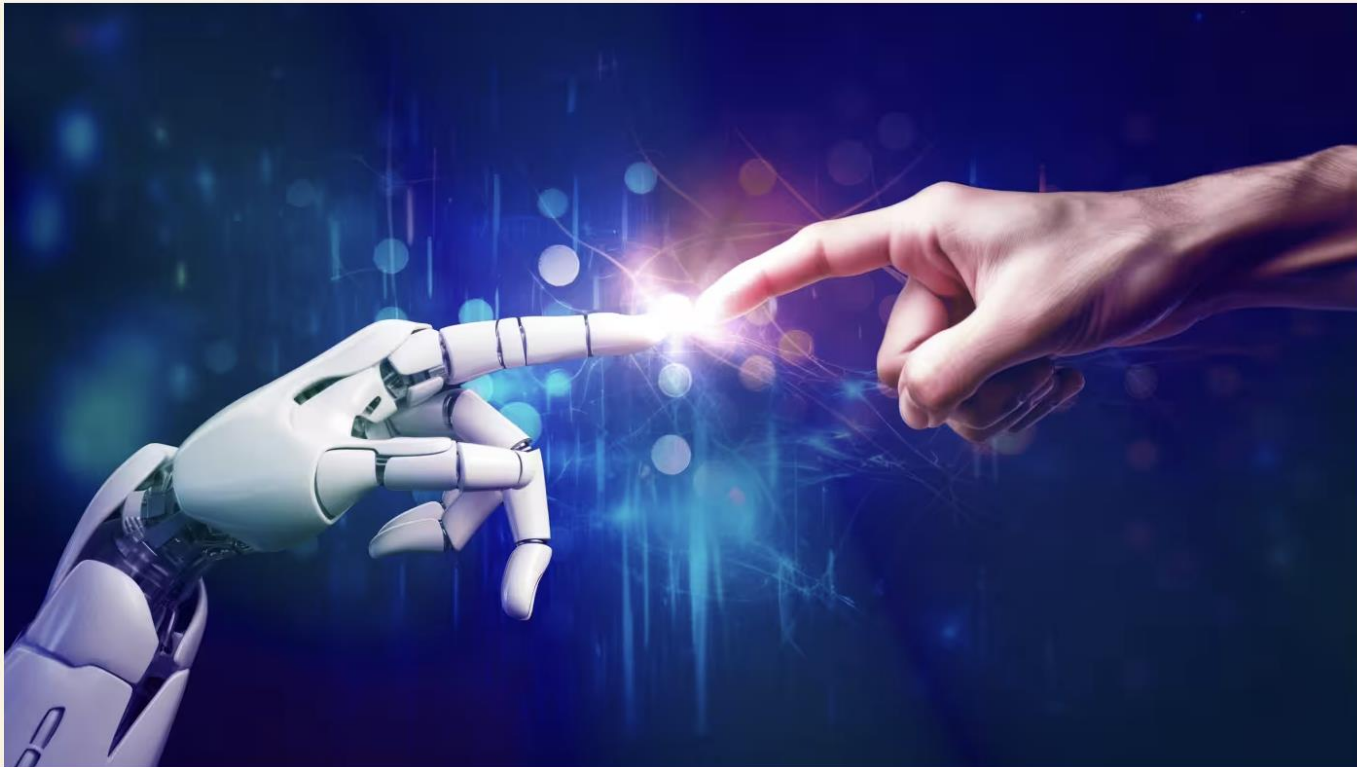
Opinion **The FT View**

AI copyright wars need a market solution

Data licensing can support a thriving ecosystem for tech firms and creators

THE EDITORIAL BOARD

+ Add to myFT



Opinion in FT
(4/3/2025)

But need of
practical measures
and a legislative fix
too!

Outline

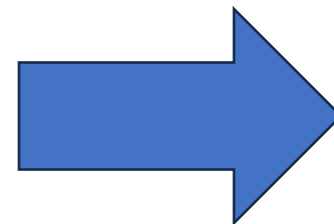
- Shifts in value creation and substitutive effects
- The risks (if no corrective measure):
 - Continuing legal uncertainties
 - AI slop
 - AI model collapse
- A bold defense of a dual system with reinforced incentives for human-generated data
 - Improved market solution
 - New unwaivable right of remuneration for individual creators

Proposal of dual approach with reinforced incentives: (1) improvement of market solution

- **Market solution based on the exclusive right:** it will work only in favor of large right owners (big press publishers, etc.), large platforms (ex. Reddit) and content aggregators
- **Improvement:** building a trusted copyright infrastructure so as to facilitate micro-licensing for the massive use of protected works online. Not only a technical issue, it includes governance aspects (with a role for CMOs for ex.)



OpenAI has been busy, and it's only halfway through the year. Credit: Mashable Composite / Ian Moore; Images / Financial Times / Associated Press / VoxMedia / News Corp / Axel Springer / The Atlantic / DotDash Meredith / Open AI / cocodava / istock / Getty

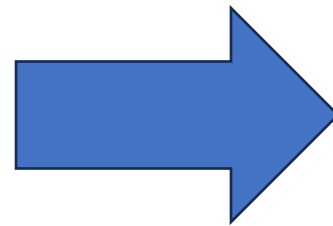


Implementing or practical measures: standardisation of right expressions & metadata (+ opt-out expressions)

Proposal of dual approach with reinforced incentives: (2) new unwaivable remuneration right for creators

- **Residual remuneration right:**

- Parallel to exclusive right
- But reserved to physical persons/creators
 - Registered as creative workers?
- Unwaivable
- Valid towards AI developers
- Could be exercised through collectives
- Triggered by conclusion of licensing deal with right owner (and/or by a presumption of use of content if the GenAI outputs are sufficiently similar)



Legislative measure: define the EU building blocks for this right of remuneration (cross-border aspects), leaving room for Member States to implement it (could be part of an ECL)

Need of linkage between the improvement of market solution and the new unwaivable remuneration right for creators

- The residual remuneration right could be considered as a ***conditional or contingent solution in the absence of an efficient and fair market solution***
- If real improvement in the distribution of revenues in favor of individual creators through the market mechanism (for ex. because CMOs or other intermediaries are involved in collecting the licensing revenues and redistributing them to individual authors, and the system becomes more efficient through the use of a copyright infrastructure), then ***a phasing-out of the remuneration right might be considered***
- This **linkage** will work as an incentive to improve the market mechanism (and to agree on standards for metadata and infrastructure)

Concluding quote

"Artificial intelligence as we know it depends entirely on a much wider set of political and social structures. And due to the capital required to build AI at scale and the ways of seeing that it optimizes, AI systems are ultimately designed to serve existing dominant interests. In this sense, artificial intelligence is a registry of power"

(Kate Crawford, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence, 2021)

KATE CRAWFORD



ATLAS OF AI

Thank you for your attention

Further information:

Alain Strowel, *ChatGPT and generative AI tools: theft of intellectual labor?*, IIC, 2023, 54, 491-494

Luc Desaunettes, Sari Depreeuw and Alain Strowel, *Study on copyright and new technologies: copyright data management and artificial intelligence*, Publications Office of the European Union, March 2022

<https://data.europa.eu/doi/10.2759/570559>
<http://hdl.handle.net.proxy.bib.ucl.ac.be/2078/266347>

Alain Strowel, Professor and attorney

alain.strowel@uclouvain.be / alain.strowel@pierstone.com