WORKING PAPER | ISSUE 09/2024 | APRIL 2024

BERTIN MARTENS

Generative arti cial intelligence (GenAI) models have stirred considerable controversy about copyright protection for AI training inputs and model outputs. e European Union's AI Act will require model developers to be transparent about their use of training inputs such as text, images and music. e EU Copyright Directive allows free text and data mining of these media inputs unless copyright holders have opted out and want license payments.

e right to opt-out amounts to economically ine cient overprotection of copyright. Free use of media content for GenAI training does not a ect media sales to consumers. Opt-outs only strengthen the bargaining position of copyright holders, who decide depending on their private interests. at generates windfall pro ts without any increase in consumer surplus or social welfare.

e licensing of training inputs reduces the quantity of data and the quality of GenAI models, creates transaction costs and reduces competition between GenAI rms. is slows down GenAI-induced innovation in media products and production processes, and productivity gains in all service sectors that apply GenAI. Ultimately, it slows down economic growth compared to what it could be with competitive and high-quality GenAI.

Bargaining over license pricing is arbitrary as there is no objective revenue benchmark to start from. Defenders of the moral right to remuneration argue (w)-7 (n)-14.90p copuenw7 (t pr)PB otenidies transction cosws dowoducts duced innoo.ts and rdis arbit69 (s AI A)9 (ctAI.)]TJ 0thdel as th

Recommended citation:
Martens, B. (2024) 'Economic arguments in favour of reducing copyright protection for generative AI inputs and outputs', Working Paper 09/2024, Bruegel

1 Introduction

Copyright is a policy tool to stimulate innovation in society. Granting exclusive private intellectual property rights, in the form of copyright, to human authors is meant to be an incentive for investment in the production of creative content such as books, music and movies. It is a tool to prevent

v e

48 . 62 - wTcommercial free-riding by others on that content in continental Europ

These properties also explain the ambiguous relationship between GenAI and the media industries. GenAI models are trained on human-produced work, including content over which artists hold exclusive copyright. At the same time, human artists use GenAI to leverage their media productivity and explore innovative new media outputs that build on and compete with existing human content. The impact of GenAI models spills over far beyond media industries to every sector in the economy.

The EU AI Act tries to settle these tensions by applying a generic formulation[2]. The draft AI Act (European Commission, 2021) and subsequent amendments by the European Parliament led to a new Article 52(c) that requires GenAI developers to respect existing copyright on training inputs, as defined in the text and data-mining exception to copyright in the Copyright in the Digital Single M[i)-m-6 (I (es(g)-4.1 (le)E6 (a)-

With GenAI, copyright issues have grown out of the confines of the creative media industries and become a wider economic issue. Finding a new GenAI-induced balance in copyright law should not only take into account the welfare of media industry producers and consumers, but of the entire economy.

On the GenAI training inputs side, we argue that granting an unconditional TDM exception for the use of copyright-protected media will not reduce the supply of creative media outputs. But it _ a^ c T b GenAI-based innovation and productivity gains, not only in media industries but across all sectors. TDM opt-outs and licensing requirements are therefore superfluous. They strengthen the bargaining position and generate windfall monopoly profits for copyright holders, at the expense of other sectors. Such opt-outs and requirements fragment the knowledge base on which GenAI models are trained, increase prices and transaction costs for training inputs, reduce competition in favour of very large G

Elkin-Koren and Weinstock (2020) argued that the EU TDM exception is too restrictive and puts EU GenAI research at a disadvantage compared to other jurisdictions that have more open and flexible provisions, including the US, Singapore, South Korea, Malaysia, Israel and Taiwan. Tyagi (2023) also recognized that the TDM exception in the EU CDSM is too narrowly defined to work for GenAI, and suggested that a broader general exception along the lines of Japan's copyright law would fit better. Article 30-4 in Japan's copyright law permits commercial TDM when not done for human enjoyment purposes (Ueno, 2021). Clearly, training of GenAI models falls in that category. However, the output of GenAI models can be used for human enjoyment. A similarly broad exception exists in the 2021 Singapore Copyright Act. It includes a mixture of a TDM exception for computational use by machine learning models (sections 243-244) and a US-style fair-use exception (sections 190-191) (Tan, 2024). However, the computational exception is subject to lawful access, including respecting paywalls and terms and conditions. The latter boils down to an opt-out possibility for content providers. The fair-use exception does not allow reproduction of the training inputs in the outputs of the AI model. Substitution with existing content remains a source of legal uncertainty. Care should be taken before jumping to conclusions about the EU's TDM disadvantage.

Margoni and Kretschmer (2022) presented a different perspective. They offered legal arguments in support of the view that there is "*no need for a TDM exception as the extraction of factual information from protected content is external to the remit of copyright*". It is only a particular expression of factual content that is protected by copyright, not the content itself. In fact, news publishers themselves, who are in the front line of TDM and copyright infringement claims against GenAI developers, widely use TDM to search continuously for new news items on the webpages of their competitors, which they then rewrite for publication on their own sites to circumvent copyright claims (Cagé *et al*, 2020). Moreover, they increasingly use GenAI services [illegible]

of scientific publishers at the time of the debate on the EU CDSM: publishers would have to incur additional costs to create machine-readable databases for TDM (Hargreaves *et al*, 2014). The AI industry has rejected that argument.

In the US, the Fair Use and Transformative Use copyright doctrines allow more freedom in the use of copyright-protected inputs for purposes other than the original purpose around which copyright holders build their business models. However, the interpretation of these doctrines is subject to uncertainty, in particular when GenAI models produce outputs that compete with the training inputs. Several copyright holders have opened court cases against GenAI firms that rely on this doctrine to access copyright-protected material for free as training inputs for their foundation models. This includes the New York Times against OpenAI[6], and Getty Images against UK-based StabilityAI[7]. The latter is particularly interesting because it revolves around alleged cross-border violations of copyright, which is essentially a territorial right. Getty Images claimed that5 (ol)15 (al5ff(ede11. (s)-2)10 (a)-1 (s)]TJ0 Tc 0 Tw 10.5

argument goes back to the continental European view of an absolute and fundamental right to a remuneration for the authors of creative content, irrespective of the economic conditions or circumstances.

2.2 The outputs side

Novelli *et al* (2024) summarised the legal questions on the outputs side as reflecting two issues: the possibility of granting copyright to outputs, and the legal relationship between the inputs used for training and the outputs. In most countries, copyright law is on the side of humans: copyright only applies to human output, not machine-

GenAI algorithms can produce creative content ranging from truly original to perfect copyright-infringing substitutes for existing human-produced

copyright-protected training inputs, without opt-out clauses and licensing conditions, increases GenAI model quality, resulting in higher profits for the AI company, higher aggregated incomes for content creators and greater consumer surplus. In short, TDM opt-outs on GenAI training data inputs give excessive protection to copyright

Steps in this direction are already being taken, for example by the French music copyright management organisation SACEM. SACEM imposed a collective opt-out from GenAI model training for all French music that falls under its authority, based on Art 4 CDSM (Spitz, 2024). As a result, French music is likely to be marginalised in GenAI models. That may have negative spillover e□ects to French artists who want to use GenAI for the production of new audiovisual materials, and to other industries that rely on French language audio and musical inputs for the production of GenAI-driven services. Whether French GenAI start-ups are able and willing to pay for licenses is not clear. GenAI producers may turn to other sources of musical inputs and marginalise typical French music in GenAI-based synthetic media. SACEM, as an intermediary in the music supply chain, might increase its own revenue from this stance, but French cultural production and society at large may lose out from this.

3.2 Pricing of copyright licenses for training inputs

Apart from quantity rationing, Art 4 CDMS also introduces monopolistic pricing for copyright licenses giving access to GenAI training inputs. Pricing of TDM licenses runs into substantial theoretical and practical problems.

From a theoretical perspective, the standard economic model of production considers that, to be fair and e□cient, remuneration of inputs should be in accordance with their marginal contribution to the value of outputs. This marginal remuneration rule is known in economics as the Euler Theorem[10]. It runs into problems when applied to GenAI models, for at least two reasons. First, unlike physical goods, data inputs and media outputs are non-rival products. They can be re-used without limits. That is the reason why copyright on these products is needed in the first place. However, non-rivalry undermines the Euler Theorem and remuneration according to marginal productivity. Doubling the volume of data inputs will not double GenAI model outputs or productivity (Romer, 1994; Buchanan and Yoon, 1999). Larger input datasets improve model performance but the marginal contribution of a single data input can vary from zero to very high, depending on the use of the model. This makes it impossible to determine an economically meaningful and e□cient licensing price for media inputs for GenAI training purposes. Pricing becomes a pure bargaining issue.

License pricing also runs into practical problems. Pricing could be an arbitrary one-o□ lump sum, irrespective of the value of the GenAI output. Pricing could also be calculated as a proportional share of the output value of GenAI models. That opens up a new set of insurmountable computational problems. At what point in the GenAI value chain should output value be used as a benchmark – the revenues of GenAI developers, or the revenues of redeployers who add additional features and datasets to models, or prices charged to final users? What about open GenAI models that are available for free? For example, OpenAI's close collaboration with Microsoft results in several implicit and explicit financial transfers between the two firms. Microsoft, in turn, builds OpenAI's models into a wide range of consumer and business services and charges for this. Whose output value should be the benchmark

[10] For a more detailed explanation, see for example *Wikipedia*, 'Euler's theorem', https://en.wikipedia.org/wiki/Euler%27s_theorem.

proteins that may have useful applications in new pharmaceutical products, a task that would have

Although prompts are considered legally insu□cient as human input to qualify for copyright on GenAI output (Hugenholtz and Quintais, 2021), complex and well-designed prompts have value in marketplaces, as the examples of Promptbase.com and the OpenAI apps store show. A complex set of prompts amounts to human-written computer code. When assembled into applications on top of a

benchmark.

References

Abbott R. and E. Rothman (2022) 'Disrupting creativity: Copyright law in the age of generative artificial intelligence', *Florida Law Review*, August

Bajari, P., V.

CMA (2023) *AI Foundation Models: Initial review*, UK Competition and Markets Authority

Dermawan, A. and P. Mezei (2023) 'The Jakarta E□ect: Could Southeast Asia's Consensus-Based Remuneration Approach Be the Blueprint for Human-Centric AI?' *Kluwer Copyright Blog*, 14 December, available at https://copyrightblog.kluweriplaw.com/2023/12/14/the-jakarta-e□ect-could-southeast-asias-consensus-based-remuneration-approach-be-the-blueprint-for-human-centric-ai/

Dreier, T. (2023) 'The challenge to attribute AI output to inputers', presentation to the European Copyright Society conference, 26 May, available at https://www.law.kuleuven.be/citip/en/citip-conferences/ecs/ecs-2023/presentations/ecs-2023-thomas-dreier.pdf

Emanuilov, I. and T. Margoni (2024) 'Memorisation in generative models and EU copyright law: an interdisciplinary view', *Kluwer Copyright Blog*, 26 March, available at https://copyrightblog.kluweriplaw.com/2024/03/26/memorisation-in-generative-models-and-eu-copyright-law-an-interdisciplinary-view/

Elkin-Koren, N. and N. Weinstock (2020) 'Transplanting Fair Use across the Globe: A Case Study Testing the Credibility of U.S. Opposition', 72 *Hastings Law Journal* 1121

Ernst and Young (2021) *Rebuilding Europe: the cultural and creative economy before and after the Covid crisis*, study commissioned by the European Grouping of Societies of Authors and Composers

European Commission (2021) 'Proposal laying down harmonized rule on artificial intelligence (the Artificial Intelligence Act)', COM/2021/206

Geiger, C. and V. Iaia (2024) 'Towards a statutory remuneration right for machine learning of GenAI', *Computer Law & Security Review* 52

Gervais, D. (2023) 'AI and CR: a reply to Matt Hervey', *Kluwer Copyright Blog*, 16 February, available at https://copyrightblog.kluweriplaw.com/2023/02/16/ai-and-copyright-a-reply-to-matt-hervey/

Hargreaves, I., L. Guibault, C. Handke, P. Valcke, B. Martens and R. Lynch (2014) *Standardisation in the area of innovation and technological development, notably in the field of Text and Data Mining*, Report from the Expert Group, European Commission, Publications O□ce of the EU

Hugenholtz, B. and J. Quintais (2021) 'Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?' *International Review of Intellectual Property and Competition Law* 52: 1190-1216

Keller, P. (2023) 'Generative AI and copyright: Convergence of opt-outs?' *Kluwer Copyright Blog*, 23 November, available at https://copyrightblog.kluweriplaw.com/2023/11/23/generative-ai-and-copyright-convergence-of-opt-outs/

Landes, W. and R. Posner (1989) '

Louie R., A. Coenen, C. Zhi Huang, M. Terry and C. Cai (2020) 'Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*: 1–13

Margoni, T. and M. Kretchmer (2022) 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology', *GRUR International* 71(8): 685-701

Nordemann J.B. (2024) 'EU law: Generative AI, copyright infringements and liability', *Kluwer Copyright Blog*, 23 January, available at https:/

Villalobos P., J. Sevilla, L. Heim, T. Besiroglu, M. Hobbhahn and A. Ho (2022) ‘Will we run out of data? An analysis of the limits of scaling datasets in Machine Learnin run

Bruegel, Rue de la Charité 33, B-1210 Brussels
(+32) 2 227 4210
info@bruegel.org
www.bruegel.org