



Kent Academic Repository

Jenkins, Georgia (2025) *A Principle of Artistic Data Sovereignty: Linking Creative Reuse to Author Remuneration*. Chicago-Kent Law Review, 100 (1). pp. 57-71. ISSN 0009-3599.

Downloaded from

<https://kar.kent.ac.uk/106556/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://scholarship.kentlaw.iit.edu/cklawreview/vol100/iss1/6>

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

A PRINCIPLE OF ARTISTIC DATA SOVEREIGNTY: LINKING CREATIVE REUSE TO AUTHOR REMUNERATION

Georgia Jenkins
University of Kent

INTRODUCTION

It would appear that creativity, as we know it, is facing extinction. Countless headlines obsess over the threat of AI-generated content to human authors as the extended copyright system grapples with the best regulatory approach. Yet, the relationship between technology and copyright law has always questioned the normative value and status of authors within the copyright system, particularly the inclusion of creative users. Today, this extends to the use of AI systems for creativity which continues to challenge the boundaries of authorial control over works.

A creativity-forward approach is vital. It starts from a foundation that authorship inherently comprises creative reuse. In the context of AI-generated content, this requires a balance within the creative process: authors must be remunerated for use of their works and access to AI systems for creativity, safeguarded. One way to structure this balance is to view authors' works, and indeed the training data input, as artistic data. This would affirm the artistic autonomy of creators, both authors and users, but also fundamentally provide a more flexible framework to ground authorial remuneration rights for creative reuse.

This framework comprises 3 parts, adopting a European lens to generative AI and creativity. Firstly, generative AI as a creative practice is located as part of a much broader dialogue of creative reuse. Secondly, the lens of artistic data sovereignty can empower authors and support new technological creative practices, specifically generative AI. Thirdly, a remuneration right should flow from artistic data training use,¹ drawing inspiration from data trusteeships. The hope is that a more comprehensive perspective on the relationship between authors' works and AI training data will forge a new outlook on the value of creativity when confronted with new technology that has the potential to enhance creative practices.

PART 1: RESTRUCTURING CREATIVITY AND AUTHORSHIP

For authors to create, they not only require autonomy and self-determination over their creations, but control over how they express themselves. Technology has improved the ability to communicate, but alongside it, once shared, works, with permission or not, can be exponentially copied and distributed online. Copyright law has responded by expanding exclusive rights, introducing value by extending control over copyright-related communications facilitated through online business models. However, copyright law is

¹ Kivanç Tatar and *et al*, 'A Shift in Artistic Practices through Artificial Intelligence' (2023) arXiv preprint arXiv:2306.10054. Accessible here: <<https://arxiv.org/abs/2306.10054>> ; Martin Senflteben and *et al*, 'Ensuring the Visibility and Accessibility of European Creative Content on the World Market: The Need for Copyright Data Improvement in the Light of New Technologies and the Opportunity Arising from Article 17 of the CDSM Directive' (2022) 13 JIPTEC 67,71 at [7].

yet to implement an approach that reflects the value of creative users' contributions. Termed creative reuse, the creativity that users partake in when they reuse copyright works, should be viewed as part and parcel of authorship.

A. The boundaries of authorship, and AI for creative reuse

Despite the challenges presented by a personality-based rationale of copyright law, there is a core component of authorship that relates to authorial control over the communication of works. In contrast to a property-based right, authorial control extends to how authors communicate both their thoughts in making a work and the work itself.² The use of authorial works, lacking authorisation, clearly impacts an author's initial communication of the work to the world. Examples of politicians using songs without authorisation for campaign-related activities spell this out quite clearly. Yet, how does this unfold in the context of creative reuse, specifically the use of authorial works for generative AI training?

If freedom of expression and art centres this discussion, it is possible to construct the boundaries of creative reuse in a manner that balances the interests of both authors and creative users, and more widely, digital creativity. European Court of Justice (CJEU) and European Court of Human Rights (ECHR) jurisprudence demonstrates an attempt to reconcile the conflicting use of freedom of expression to support both authors and users.³ For example, the *Deckmyn* judgement requires a proportionality assessment that includes the legitimate interest of the copyright holder to not be associated with a harmful message when assessing the limits of parody.⁴ National judgements have also introduced a more practical approach which centres upon the degree to which a parody of the work would conflict with the intended use of the work, and its purpose to support freedom of expression.⁵

However, in specific cases of parody, and potentially pastiche, the allegedly infringing work is clear. Conversely, when authorial works are used to train generative AI, the output largely remains unclear until its actual use. A personality-based authorial claim to control works to train generative AI necessitates 'artistically' harmful output as well as transparency over artistic data training sets. The piecemeal character of lawfully accepted creative reuse, particularly in the EU, additionally lacks the flexibility to support the use of generative AI as a creative practice.⁶ Instead, a creativity-forward approach would extend beyond individual claims to control artistic communication and acknowledge the cumulative and collaborative nature of creative communities.

Building from freedom of expression and art, a modern approach to authorial consent calls for a mutual duty of creativity among authors and creative users. One that allows

² Neil Netanel, 'Copyright Alienability Restrictions and the Enhancement of Author Autonomy: A Normative Evaluation' (1993) 24 Rutgers LJ 347, 374.

³ 2018 CA-Versailles, 16 March 2018, RG 15/06029 – *Malka v Klasen*; Alan Hui and Frédéric Döhl, 'Collateral Damage: Reuse in the Arts and the New Role of Quotation Provisions in Countries with Free Use Provisions after the ECJ's *Pelham*, *Funke Medien* and *Spiegel Online* Judgments' (2021) 52 IIC 852.

⁴ *Johan Deckmyn and Vrijheidsfonds VZW v Helena Vandersteen and Others* (Case 201/13) [2014], ECLI:EU:C:2014:2132.

⁵ Cass. Civ, 22 May 2019, no. 18-12718 - *Le Point*; BGH, GRUR 2016, 1157 – *Auf fett getrimmt* (Fat cropped); OLG Jena, 2015, 2 U 674/14 – *Helene Fischer*.

⁶ Eleonora Rosati, 'Copyright in the EU: in search of (in)flexibilities' (2014) 9 JIPLP 585.

creative reuse, shielding it from exclusive rights to the extent that authors' interests over the artistic communication are reconciled. This would largely remove the influence and impact of non-authorial rights over works which control creative reuse based on an investment rationale. It has the potential to reshape and restructure the copyright system specific to digital creativity in a way that truly responds to creative interests, including generative AI art.

Termed, anticipatory aesthetics, neural media mediates and challenges traditional artistic practices by disrupting authorial intention.⁷ When generative AI is viewed as a creative practice, authorial intent is disrupted, reshaped and significantly, "re-presented" as what output comprises, or not, is the result of an opaque relation between "autographic and the algorithmic".⁸ While this brings questions of authorship to the fore, it does not entirely negate the creative nature of the reuse. Indeed, others have certainly called for a fundamental restructuring of copyright law, in response to generative AI, premised on reimagining cultural ownership, creativity and creative expression.⁹

A creativity-forward approach should draw on the participatory nature of authorship, the use of generative AI as a creative practice, and the mediating influence of AI within these practices to balance the interests of authors and creative users.

B. Permissioned culture, creativity and AI-generated content

Even with the promise that a more balanced approach to generative AI could bring creativity online, it appears that the permissioned culture which characterises copyright law generally will regulate creative generative AI. Whether referential or transformative, generative AI inherently conflicts with commercial exploitation rights.¹⁰ As seen through the protection of technological protection measures, 'content datafication'¹¹, alongside entrenched contractual bargaining positions of authors and big tech companies, generative AI presents similar issues that lead to increasing proprietisation of creative expression through the copyright system.

The introduction of the text and data mining (TDM) exceptions in the EU Copyright Digital Single Market (DSM) Directive confirms as much. Its legislative nature almost disguises the expansion of the reproduction right without actual amendment.¹² While Article 3 CDSM allows TDM specific to research, it is more likely that creative reuse would fall

⁷ Suk Kyoung Choi, Steve DiPaola and Hannu Töyrylä, 'Artistic Style Meets Artificial Intelligence' (2021) 4 *Journal of Perceptual Imaging* 020501-1. Choi and *et al* explain that neural media refers to technologies used in "AI art", derived from machine learning and deep artificial networks.

⁸ Suk Kyoung Choi and Steve DiPaola, 'AI as *other*: An art-as-research approach to generative AI art practice' (2023) *Proceedings of the International Conference on Computational Creativity (ICCC)*. Available here: < https://computationalcreativity.net/iccc23/papers/ICCC-2023_paper_109.pdf>. Choi and DiPaola reflect on the striking resemblance between the black box of generative AI alongside traditional artistic outputs within aesthetics.

⁹ Martin Zeilinger, *Tactical Entanglements* (Meson Press 2021) p. 173.

¹⁰ Lawrence Lessig, *Free culture: How big media uses technology and the law to lock down culture and control creativity* (Penguin 2014) 254-9.

¹¹ Valentina Moscon, 'Data Access Rules, Copyright and Technological Protection Measures in the EU' in Florent Thouvenin and *et al*, *Kreation Innovation Märkte - Creation Innovation Markets: Festschrift Reto M. Hilty* (Springer 2024) p. 1033

¹² *Ibid* 1041.

outside these parameters. However, it may benefit from a more general exception pursuant to Article 4 if the rights holder has not reserved the right to extraction and the extraction is carried out by a person that has “lawful access”.¹³ Rights holders can reserve their TDM related rights by opting out which should increase their bargaining position with AI developers while simultaneously entrenching an asymmetrical power imbalance between creators and dominant AI companies.¹⁴

This ‘backdoor’ propertisation of data through the copyright system is reminiscent of the introduction of TPMs. This somewhat denies the significance of the idea/expression dichotomy of copyright protection, and by extension the claim that mere facts and data should fall outside the copyright system. The copyright system has long been used as a doctrinal vehicle for securing and entrenching exploitative rights specific to new online business models. There seems to be a growing call to restrict data ownership within the copyright system. While most focus on the detrimental impact that data propertisation will have on AI innovation,¹⁵ there is also a clear need to consider the position of authors and users, not merely AI developers.

It is well established that generative AI will detrimentally impact authors, eventually replacing them.¹⁶ One solution is to focus on the authenticity factor of human creativity, which, to an extent, copyright law often fails to protect.¹⁷ The social and cultural value of human creativity could be recognised within the copyright system by attributing authorial rights. As AI systems by their very nature require access to works to train, there is a significant causative role that must be recognised by law. While authors’ rights and personality-based justifications may provide a good foundation to centre this discussion, ultimately these rights should be formulated on a community-level basis.

Without this baseline, access to and use of authorial works will be set contractually and subject to TPMs. A creative community-led call for authorial rights specific to training data input is the first step. As the EU “move[s] towards a process of propertisation of non-personal data”¹⁸, there needs to be a structure set in place that ensures that creativity, including generative AI as a creative practice, is safeguarded. Not merely an extension of the argument to incentivise AI system innovation. Perhaps the EU is set to introduce a further amendment, if public outlash and copyright clairvoyants are correct.¹⁹

¹³ CDSM Directive.

¹⁴ Gina Maria Ziaja, ‘The Text and Data Mining Opt-out in Article 4(3) CDSMD: Adequate Veto Right for Rightsholders or a Suffocating Blanket for European Artificial Intelligence Innovations’ (2024) JIPLP.

¹⁵ Thomas Margoni and Martin Kretschmer, ‘A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology’ (2022) 71 GRUR International 685, 686.

¹⁶ Dan L. Burk, ‘Cheap Creativity and What It Will Do’ (2023) 57 Georgia Law Review 1669, 1680; Mark A. Lemley, ‘How Generative AI Turns Copyright Upside Down’ (2024 forthcoming) Colum. Sci. & Tech. L. Rev.; Geiger [n 25].

¹⁷ Burk [n 16] 1682, 1688 & 1669.

¹⁸ Margoni and Kretschmer [n 14] 699.

¹⁹ Christophe Geiger, ‘When the robots (try to) take over: of artificial intelligence, authors, creativity and copyright protection’ in in Florent Thouvenin and *et al*, *Kreation Innovation Märkte - Creation Innovation Markets: Festschrift Reto M. Hilty* (Springer 2024) p. 77. Geiger suggests that following the AI Act transparency provisions, intense public debate on authorial works used for training AI models, that new amendment is forthcoming.

PART 2: EMPOWERING AUTHORS AND FOSTERING CREATIVITY ONLINE

So, how best to structure the fundamental right authors to control the use of their works as training data for AI models? The discussion should be grounded through the lens of data autonomy to unpack the social and cultural value of human creativity. Termed artistic data, it provides the legal vehicle to build a limitation-based remuneration right that is creativity-centred, specific to authors, while ensuring space for generative AI as a creative practice.

A. Viewing creativity as artistic data

When we refer to ‘access’ to works in copyright law, what does this mean specific to training data input? Simply put, generative AI works through a process called diffusion. This involves using large data sets to train AI. These data sets undoubtedly comprise authorial works. In the case of the non-profit Large-scale Artificial Intelligence Open Network (LAION), used by Stability AI to train its text-to-image generator, a web crawler scrapes the internet, making temporary reproductions of retrieved images from URLs.²⁰ These images are then analysed and evaluated in a pre-training process where the raw data is converted into a suitable format to train the AI model.²¹

One way of assessing whether LAION contains unauthorised images scraped from the internet, is the website ‘Have I Been Trained?’ as the data set is open source. This is exactly what prompted German photographer, Robert Kneschke, to send a cease-and-desist letter in 2023 to LAION, requiring that his images were removed from the data set. LAION replied that there was nothing to delete, and returned the favour with a demand for €887 EUR for an unjustified copyright claim. The Hamburg Regional Court is set to decide the matter in late April 2024. Some have speculated on the impact of the newly minted TDM exceptions, particularly the timing of Kneschke’s opt-out.²²

Regardless of the outcome, it appears that LAION’s access to, and use of works is difficult to reconcile within the copyright system. Sure, the TDM exceptions play a role in fleshing out boundaries of lawful behaviour in the wake AI, but as always in copyright law, it is a response which moulds the extent of exclusive rights in the digital environment. One thing remains certain, supported largely by the public dissent of generative AI trained on authorial works, that the latter has an authentic human value that plays a significant causative role in the development of AI training data sets which is left unaccounted. But if the use made of authorial works fails to squarely fit within current boundaries of the copyright system, how should it be restructured to reflect this value?

Artistic data is a preferable concept to encompass the data preparation process, encompassing scraping and transformation of the work into raw data. When human creativity is viewed as artistic data it provides an opportunity to untie the copyright

²⁰ Andres Guadamuz, ‘A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Output’ (Forthcoming 2024) GRUR International.

²¹ Ibid; Rui-Jie Yew, ‘Break It ‘Til You Make It. An exploration of the Ramifications of Copyright Liability Under a Pre-training Paradigm of AI Development’ (2024 Symposium on Computer Science and Law) 62

²² Andres Guadamuz, ‘Photographer sues LAION for copyright infringement’ (TechnoLlama, 5 May 2023) <<https://www.technollama.co.uk/photographer-sues-laion-for-copyright-infringement>> [Accessed 5 April 2024].

Gordian knot of where access is often sacrificed at the behest of extending copyright protection online. Instead of forcing provisions implemented in response to other communication means, artistic data would respond specifically to inherent human value that is siphoned by data set providers and relied upon to train AI systems.

Ideally artistic data would sit outside of personal data requirements.²³ Construed as raw non-personal data,²⁴ it refers to data generated from digitally accessible authorial works. Generally, trade secrets and confidential information are best placed to assert ownership claims over unstructured raw data, however it comes with obligations to ensure its secrecy. This is difficult to apply to artistic works which require communication where the raw data can be ascertained.

Further while some generative AI systems disclose which data sets they use, others do not, such as Midjourney. Though the EU AI Act requires transparency,²⁵ the cumulative worth of artistic data in training AI models should not result in an individual claim of training. Instead, a community-based assessment would assume both the inherent social and cultural value in artistic data, leading to the assumption that generative AI relies on the authenticity of human creativity.

B. Artistic data sovereignty, a guiding principle

There is a steady call throughout EU commentary to respond to the challenges presented by AI systems with a limitation-based remuneration right to regulate TDM of copyright works.²⁶ However, before delineating the boundaries of this limitation, there should be a theoretical foundation to support these claims. If creativity is viewed as artistic data, a limitation-based remuneration right could be constructed as a unique claim of data sovereignty. One that affirms the value of human authenticity in creativity, but equally, provides sufficient space for generative AI as a creative practice.

A traditional personality-based approach lays the foundation to consider creativity as an extension of the author and an execution of their will.²⁷ How authors communicate creative processes and the works that follow, could be framed as fundamental elements human authenticity. Any unauthorised use, of course, infringes the very meaning of authorship. However, there are limits to authorial control over creative reuse, particularly the idea that authorial will must coexist with other creative users.²⁸ In the

²³ Though perhaps it could also be categorized as behaviour-generated data? Karl-Heinz Fezer, 'Data Ownership of the People. An intrinsic Intellectual Property Law Sui Generis Regarding People's Behaviour-generated Informational Data' (2017) ZGE/IPJ 356.

²⁴ Tommaso Fia, 'An Alternative to Data Ownership: Managing Access to Non-Personal Data through the Commons' (2020) 21 Global Jurist

²⁵ Article 4.

²⁶ Geiger [n 19]; Martin Senftleben, 'Generative AI and Author Remuneration' (2023) 54 IIC 1535; Giancarlo Frosio, 'Should We Ban Generative AI, Incentivise It or Make It a Medium for Inclusive Creativity?' in Enrico Bonadio and Caterina Sganga (eds) *A Research Agenda for EU Copyright Law* (Elgar forthcoming) Accessible here https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4527461 [Accessed 10 April 2024].

²⁷ Neil Netanel, 'Copyright Alienability Restrictions and the Enhancement of Author Autonomy: A Normative Evaluation' (1993) 24 Rutgers LJ 347, 374.

²⁸ See, Amy Lai, *The Right to Parody: Comparative Analysis of Copyright and Free Speech* (Cambridge University Press 2019).

context of training AI systems, particularly data preparation, there is a strong authorial claim to control the unauthorised inclusion of works. There is no way for authors to know the context of the data set, nor how it will be used to train an AI model, and lastly, the output.²⁹ They stand to lose some cognitive sovereignty over future artistic practices and creative reuses of their works.³⁰

The notion of authorial data autonomy is based on this uncertainty and provides a starting point to assess and determine the boundaries of access to and use of authorial works for the purposes of training AI. This is nothing new. Constitutional personality rights already extend to protect informational self-determination.³¹ While normally a claim-based in moral rights, a fresh outlook on personhood theory specific to human creativity used to train AI systems could provide the structure to reflect creative practices including reuse more practically.³²

Building a call for artistic data sovereignty from a personhood foundation is helpful given the unwaivable nature of personality rights. It also recognizes the value of human authenticity in creativity that is necessary component to train any generative AI system. But how best to structure this principle? In the context of behaviour-generated data, some have suggested a sui generis IP right to support individual autonomy, by allowing a property-based right to be claimed.³³ However, the focus should not be on ownership rights, but setting the boundaries of access specific to creativity,³⁴ given the implications of data ownership³⁵.

When generative AI is used as creative practice, it both diminishes and enhances freedom of expression. As authors' intentions are mediated and reshaped by AI, anticipatory aesthetics suggests that:

*“[A]rtistic expression as an evolving process of self-apprehension leaves in its wake a data trace—a praxis narrative of affective intent in the multimodal ecology of creative practice—from which AI might learn about and reflexively extend human anticipatory acts”.*³⁶

²⁹ Choi and DiPaola [n 6] 1.

³⁰ Ulrich Beck, *Risk Society: Towards a New Modernity* (London: Sage Publications, 1992) pp.53-54.

³¹ Johannes Eichenhofer, 'The Constitutional Protection of Images' in Thomas Dreier and Tiziana Andina (eds) *Digital Ethics – The Issue of Images* (Nomos 2022).

³² Christopher S. Yoo, 'Self-Actualization and the Need to Create as a Limit on Copyright' (2021) Faculty Scholarship at Penn Law 2041.

³³ Karl-Heinz Fezer, 'Data Ownership of the People. An Intrinsic Intellectual Property Law Sui Generis Regarding People's Behaviour-generated Informational Data' (2017) 9 German Intellectual Property Journal 356.

³⁴ Ingrid Schneider, 'Digital Sovereignty and Governance in the Data Economy: Data Trusteeship Instead of Property Rights on Data' in Christine Godt and Matthias Laming (eds) *A Critical Mind* (Springer 2023) pp. 376-7; P. Bernt Hugenholtz, 'Data property: Unwelcome Guest in the House of IP' in H. Ullrich and et al, *Kritika: Essays on Intellectual Property* (Elgar 2018); Andreas Wiebe, 'Protection of industrial data – a new property right for the digital economy?' (2017) 12 JIPLP 62.

³⁵ Fia [n 24] 183.

³⁶ Choi and DiPaola [n 7] 2.

The interaction between the autographic and the algorithmic hinges on the author's self-realization through art which improves the AI system's ability to generate content, when used as an artistic practice. The characterisation of the relationship between artist and AI as a *data trace* strengthens the author's right to control artistic data relating to their works. Conversely, the mediating influence of AI systems, when used as an artistic practice, also bolsters the position of creative users who play a significant role in the reshaping and re-presentation of creativity.³⁷

Distinct from a new IP right, artistic data sovereignty should comprise two clear elements: consent and remuneration. Copyright law often adopts an implied licence framework to structure lawful uses. Indeed, a more positive outlook on the TDM exception would characterize it as just that, a law-based authorisation. However, as hinted earlier, it is granted without any affirmation of the value of human authenticity in creativity. A more balanced approach would recognize a principle of artistic data sovereignty, that alongside a law-based authorisation requires remuneration.

PART 3: ARTISTIC DATA SOVEREIGNTY AND AUTHORIAL REMUNERATION

If authors are to be paid based on the value of their creative authenticity, termed artistic data, how should this remuneration right be structured within the copyright system? The EU has recently implemented obligations to drive platform licensing,³⁸ which in some member states, has resulted in a direct line of payment between the platform and authors.³⁹ Collective licensing is used to manage the new lines of remuneration to authors through CMOs,⁴⁰ while simultaneously allowing creative reuses (parody, pastiche, and caricature). The sentiment is clear – a paid and permitted approach best balances copyright interests. Yet is this approach the right fit for copyright works used to train AI models? In line with viewing human creative authenticity through the lens of artistic data, it is worth exploring options that could more aptly reflect the access to, and transformation of, artistic data, during the data preparation process.

A. An artistic data trusteeship?

While some advocate for a communitarian model to facilitate a digital commons,⁴¹ a data trusteeship may provide a more suitable structure given their flexibility for balancing property-based interests with access. Perhaps most well-known is the Biotrust Model which regulates both property rights, rights reservation, access and funding relating to

³⁷ Ibid 1.

³⁸ 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 [2019] OJ L 130/92

³⁹ See Germany's implementation of article 17 DSM Directive. Urheberrechts-Diensteanbieter-Gesetz vom 31. Mai 2021 (BGBl. I S. 1204, 1215) (Act on the Copyright Liability of Online Content Sharing Service Providers)

⁴⁰ Mattias Leistner, 'The Implementation of Art. 17 DSM Directive in Germany – A Primer with Some Comparative Remarks' (2022) 71 GRUR International 909, 920.

⁴¹ Alan Chan, Herbie Bradley and Nitarshan Rajkumar, 'Reclaiming the Digital Commons: A Public Data Trust for Training Data' (2023) arXiv:2303.09001; more broadly, Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (CUP 2015).

genomic biobanks.⁴² Similarly to biobanks, the use of authorial works to train generative AI brings ethical questions to the fore that relate to the very definition of human nature. While biobanks comprise large data sets containing genetic data,⁴³ AI training data sets, such as LAION, contain artistic data relating to human creative authenticity. Though commentary has identified non-for-profit data trusts as a flexible solution to regulate sensitive data, particularly to reconcile “power asymmetries”⁴⁴, they depend on a context-based assessment⁴⁵ which lacks sufficient discussion in relation to artistic data and copyright.

Not-for-profit data trusteeships operate on two criteria: (1) an independent third party exercises sovereign rights over data, and (2) this third party is delegated control by data donors.⁴⁶ The data trustee, a legal fiduciary, is responsible for the property and the beneficiary is the public.⁴⁷ This structure provides a foundation to appropriately balance ethical, legal and societal values specific to the subject matter of the data. In the context of artistic data, human creative authenticity and its cultural significance would be driving values that the data trustee must safeguard. Additionally, and perhaps notably, not-for-profit data trusts must be “trustworthy, professional and, above all, legally and financially independent”⁴⁸.⁴⁹

In the EU, the adoption of data trusts aligns with the EU data strategy which foresees the creation of Common European Data Spaces to support data sovereignty.⁵⁰ The Fair MusE project illustrates a good starting point.⁵¹ By evaluating the ‘fairness’ of algorithms on music streaming providers, it promotes greater transparency relating to data collection and its use. Data spaces, based on the concept of data altruism, also require voluntarily shared data generally without remuneration, though compensation is possible for data reuse.⁵² Though it is unclear whether artistic data should fall under the umbrella of data

⁴² David E. Winkhoff & Larissa B. Neumann, ‘Towards a Social Contract for Genomics: Property and the Public in the ‘Biotrust’ Model’ (2005) 1 *Genomics, Society and Policy* 8.

⁴³ Clare Bycroft and *et al*, ‘The UK Biobank resource with deep phenotyping and genomic data’ (2018) 562 *Nature* 203.

⁴⁴ Ingrid Schneider, ‘Data Stewardship by Data Trusts: A Promising Model for the Governance of the Data Economy’ in Caludia Padovani and *et al*, *Global Communication Governance at the Crossroads* (Springer 2024) p. 333.

⁴⁵ Sylvie Delacroix and Neil D. Lawrence, ‘Bottom-up data Trusts: disturbing the ‘one size fits all’ approach to data governance’ (2019) 9 *International Data Privacy Law* 236.

⁴⁶ Schneider [n 44] 393.

⁴⁷ *Ibid* 387.

⁴⁸ *Ibid*.

⁴⁹ Particularly given the pitfalls of collective licensing. See, Rebecca Tushnet, ‘Comment on “Frontiers: The Interplay of User-Generated Content, Content Industry Revenues, and Platform Regulation: Quasi-Experimental Evidence from YouTube’ (2023) 43 *Marketing Science* 1; Jonathan Band, ‘Cautionary Tales About Collective Rights Organizations’ (2012) 21 *International Law Review* 687.

⁵⁰ Commission, ‘A European strategy for data’ COM (2020) 66 final.

⁵¹ Guiseppe Mazziotti and Heritiana Ranaivoson, ‘Can Online Music Platforms Be Fair? An Interdisciplinary Research Manifesto’ (2024) 55 *IIC* 249.

⁵² Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) recital 45, Article 2(16); Vikas Kathuria, ‘Comparative Assessment of Non-Personal Data Access Frameworks’ in *Technologies, Law and Society* (Edinburgh University Press 2024, forthcoming) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4746707> [Accessed 9 April 2024].

spaces, subject to independent stewardship requirements, not-for-profit data trusts provide a helpful starting point to explore balancing mechanisms specific to artistic data.

They have the potential to bolster data agency,⁵³ a missing factor when considering the inclusion of authorial works in data sets used to train generative AI, but also provide a good foundation to consider the boundaries for mitigating access to data for the purposes of creative reuse.

B. Artistic data nexus to the copyright system

By viewing the impact of training generative AI through the lens of artistic data, a more direct line of remuneration can be made for authors. Currently in the EU, the focus appears to be on the extent to which copyright works used to train AI infringe copyright and related rights, particularly, a *sui generis* database right.⁵⁴ It is important to note that the database right is not an authorial right but based on an investment rationale for the development of an “information market”.⁵⁵ Commentary explains that its enforcement, particularly in the UK, has evidenced widespread parasitism where ‘broad notions of unfairness’ are used only when income is at risk.⁵⁶

Instead, artistic data purely focuses on authors’ rights and the value of human creative authenticity. It responds to the age-old problem: the weak bargaining position of authors when negotiating exploitation.⁵⁷ In this sense, when we question the liability of training AI systems using authorial works, we really question (in vain) the normative status of authors within culture and creativity. Indeed, the introduction of generative AI systems appears to only be the latest emerging technology in a long list that have displaced authors. Yet, there is a change.

While previously these technologies linked to dissemination of works, generative AI risks not simply authorial control over communication, but the generation of creativity,⁵⁸ itself. This difference deepens the threat against human creative authenticity and necessitates a fundamental rethinking of authors’ rights and data. One example are the equitable remuneration principles outlined in the EU Digital Single Market (DSM) Directive. Here, member states are tasked with data transparency relating to the exploitation of works and performances to whom they have licenced or transferred their rights,⁵⁹ as well as all revenues generated and remuneration due.⁶⁰ Any disparity also affords an authorial right to amend contractual remuneration.⁶¹

⁵³ Schneider [n 44] 397.

⁵⁴ Philipp Hacker, ‘A legal framework for AI training data – from first principles to the Artificial Intelligence Act’ (2021) 13 Law, Innovation and Technology 257, 277; Margoni and Kretschmer [n 22] 2.

⁵⁵ Guido Westkamp, ‘Data(base) Rights? – Misappropriation, Property, and Tales of Trials and Tribulations’ in Sharon Sandeen, Ansgar Ohly & Christopher Rademacher (eds) *Research Handbook on Information Law and Governance* (Elgar 2020) p. 99; citing Case C-46/02, *Fixtures Mktg. v Oy Veikkaus Ab.* [2004] ECR I-10365, para 33.

⁵⁶ Westkamp [n 53] 99.

⁵⁷ Jane C. Ginsburg, ‘Authors’ remuneration: reforms to wish for’ in Gustavo Ghidini and Valeria Falce (eds) *Reforming Intellectual Property* (Elgar 2022) p. 122.

⁵⁸ Burk [n 16].

⁵⁹ DSM Directive, recital 76.

⁶⁰ *Ibid* article 19.

⁶¹ *Ibid* article 20.

Yet, despite these benevolent aims, the focus on fair contracts, has not extended to creative users.⁶² And the TDM exceptions undoubtedly risk generative AI as a creative practice due to either the scientific research limitation or potential reservation of rights. There is also considerable tension between the copyright acts infringed and the process of preparing data to train generative AI, making it difficult to practically reconcile TDM within the copyright system. Any further legislative or judicial intervention, codifying TDM as an infringing act, lacks the nuance needed to structure and bolster access to artistic data.

Drawing from a data trusteeship, any legislative intervention should first be premised on access, not rights. Further, if creativity is to be valued, the rationale should not be based on an investment rationale such as to incentivise the development of AI systems. Any form of authorial remuneration should be steeped in the social and cultural benefits of creativity which includes creative users as authors. It removes the layer of complexity which non-authorial rights, such as a *sui generis* database holder, may enforce only in cases where income is affected, and places the focus on fostering creativity online. Lastly, as even a direct line of remuneration between author and the AI training data set provider, requires a limitation to support access to artistic data, this would need to comply with the three-step test.

Whether this takes forms as a statutory licence or levy system, there is a real need to apply not-for-profit data trusteeship principles to those tasked with facilitating payments.

CONCLUSION

This article began with the premise that creativity is soon to be extinct. Policy and legislative inventions in the EU have attempted to grapple with the impact of generative AI by introducing a TDM exception to balance rights holders and AI developers. However, this solution does not directly support authors. Instead, it refers to rights holders which appears to be more focused on incentivising the development of AI systems, not supporting authors, and more widely, valuing human creativity.

Considering these developments, it is suggested that a serious rethinking of the normative status of authors both within and outside the copyright system is required. The article proposes that a more effective to value human creative effort is through the concept of artistic data. The concept more closely aligns with the actual process of data preparation which transforms and incorporates authorial works into data sets to train generative AI. It also is based on the premise that human creative authenticity is a significant and causative factor of generative AI output and strengthens authorial data autonomy.

When reflecting on calls within commentary to implement a limitation-based remuneration right, the artistic data sovereignty is a helpful starting point. Data trust principles can ensure that a structure is used that safeguards creativity, both for authors and creative users by requiring access and remuneration for reuse, as opposed to granting ownership rights. This theoretical shift is significant to provide a proportional

⁶² Jens Schovsbo, 'Making sure copyright works – safeguarding authors and users' rights' in Gustavo Ghidini and Valeria Falce (eds) *Reforming Intellectual Property* (Elgar 2022) p. 211.

balance between initial authorial works, and circumstances where generative AI is used as a creative practice. The streamlined remuneration right, between authors and those responsible for AI training data sets, also supports its theoretical base, creativity, because it is linked to ensuring access for these purposes, not an investment rationale.

REFERENCES

- Band, J., 'Cautionary Tales About Collective Rights Organizations' (2012) 21 *International Law Review* 687
- Beck, U., *Risk Society: Towards a New Modernity* (London: Sage Publications, 1992)
- Burk, D. L., 'Cheap Creativity and What It Will Do' (2023) 57 *Georgia Law Review* 1669
- Bycroft, C., and *et al*, 'The UK Biobank resource with deep phenotyping and genomic data' (2018) 562 *Nature* 203
- Chan, A., and *et al* 'Reclaiming the Digital Commons: A Public Data Trust for Training Data' (2023) arXiv:2303.09001
- Choi, S.K., and *et al*, 'AI as *other*: An art-as-research approach to generative AI art practice' (2023) *Proceedings of the International Conference on Computational Creativity (ICCC)*. Available here:<https://computationalcreativity.net/iccc23/papers/ICCC-2023_paper_109.pdf> — and *et al*, 'Artistic Style Meets Artificial Intelligence' (2021) 4 *Journal of Perceptual Imaging* 020501-1
- Delacroix, S., and Lawrence, N. L., 'Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance' (2019) 9 *International Data Privacy Law* 236
- Eichenhofer, J., 'The Constitutional Protection of Images' in Thomas Dreier and Tiziana Andina (eds) *Digital Ethics – The Issue of Images* (Nomos 2022)
- Fezer, K. H., 'Data Ownership of the People. An intrinsic Intellectual Property Law Sui Generis Regarding People's Behaviour-generated Informational Data' (2017) *ZGE/IPJ* 356
- Fia, T., 'An Alternative to Data Ownership: Managing Access to Non-Personal Data through the Commons' (2020) 21 *Global Jurist*
- Frosio, G., 'Should We Ban Generative AI, Incentivise It or Make It a Medium for Inclusive Creativity?' in Enrico Bonadio and Caterina Sganga (eds) *A Research Agenda for EU Copyright Law* (Elgar forthcoming) Accessible here <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4527461> [Accessed 10 April 2024].

Geiger, C., 'When the robots (try to) take over: of artificial intelligence, authors, creativity and copyright protection' in Florent Thouvenin and *et al*, *Kreation Innovation Märkte - Creation Innovation Markets: Festschrift Reto M. Hilty* (Springer 2024)

Ginsburg, J. C., 'Authors' remuneration: reforms to wish for' in Gustavo Ghidini and Valeria Falce (eds) *Reforming Intellectual Property* (Elgar 2022)

Guadamuz, A., 'Photographer sues LAION for copyright infringement' (TechnoLlama, 5 May 2023) <<https://www.technollama.co.uk/photographer-sues-laion-for-copyright-infringement>> [Accessed 5 April 2024].

——, 'A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Output' (Forthcoming 2024) GRUR International.

Hacker, P., 'A legal framework for AI training data – from first principles to the Artificial Intelligence Act' (2021) 13 *Law, Innovation and Technology* 257

Hughenholz, 'Data property: Unwelcome Guest in the House of IP' in H. Ullrich and *et al*, *Kritika: Essays on Intellectual Property* (Elgar 2018)

Hui, A., and Döhl, F., 'Collateral Damage: Reuse in the Arts and the New Role of Quotation Provisions in Countries with Free Use Provisions after the ECJ's Pelham, Funke Medien and Spiegel Online Judgments' (2021) 52 *IIC* 852

Kathuria, V., 'Comparative Assessment of Non-Personal Data Access Frameworks' in *Technologies, Law and Society* (Edinburgh University Press 2024, forthcoming) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4746707> [Accessed 9 April 2024]

Lai, A., *The Right to Parody: Comparative Analysis of Copyright and Free Speech* (Cambridge University Press 2019)

Leistner, M., 'The Implementation of Art. 17 DSM Directive in Germany – A Primer with Some Comparative Remarks' (2022) 71 *GRUR International* 909, 920.

Lemley, M.A., 'How Generative AI Turns Copyright Upside Down' (2024 forthcoming) *Colum. Sci. & Tech. L. Rev.*

Lessig, L., *Free culture: How big media uses technology and the law to lock down culture and control creativity* (Penguin 2014)

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

Mazziotti, G., and Ranaivoson, H., 'Can Online Music Platforms Be Fair? An Interdisciplinary Research Manifesto' (2024) 55 *IIC* 249

Merges, R., *Justifying Intellectual Property* (Harvard University Press 2011).

Moscon, V., 'Data Access Rules, Copyright and Technological Protection Measures in the EU' in Florent Thouvenin and *et al*, *Kreation Innovation Märkte - Creation Innovation Markets: Festschrift Reto M. Hilty* (Springer 2024)

Netanel, N., 'Copyright Alienability Restrictions and the Enhancement of Author Autonomy: A Normative Evaluation' (1993) 24 Rutgers LJ 347

Ostrom, E., *Governing the Commons: The Evolution of Institutions for Collective Action* (CUP 2015)

Rosati, E., 'Copyright in the EU: in search of (in)flexibilities' (2014) 9 JIPLP 585.

Schneider, I., 'Digital Sovereignty and Governance in the Data Economy: Data Trusteeship Instead of Property Rights on Data' in Christine Godt and Matthias Laming (eds) *A Critical Mind* (Springer 2023)

—— 'Data Stewardship by Data Trusts: A Promising Model for the Governance of the Data Economy' in Caludia Padovani and *et al*, *Global Communication Governance at the Crossroads* (Springer 2024)

Schovsbo, J., 'Making sure copyright works – safeguarding authors and users' rights' in Gustavo Ghidini and Valeria Falce (eds) *Reforming Intellectual Property* (Elgar 2022)

Senflteben, M., and *et al*, 'Ensuring the Visibility and Accessibility of European Creative Content on the World Market: The Need for Copyright Data Improvement in the Light of New Technologies and the Opportunity Arising from Article 17 of the CDSM Directive' (2022) 13 JIPEEC 67

—— 'Generative AI and Author Remuneration' (2023) 54 IIC 1535.

Tatar, K., and *et al*, 'A Shift in Artistic Practices through Artificial Intelligence' (2023) arXiv preprint arXiv:2306.10054. Accessible here: <<https://arxiv.org/abs/2306.10054>>

Tushnet, R., 'Comment on "Frontiers: The Interplay of User-Generated Content, Content Industry Revenues, and Platform Regulation: Quasi-Experimental Evidence from YouTube"' (2023) 43 Marketing Science 1

Westkamp, G., 'Data(base) Rights? – Misappropriation, Property, and Tales of Trials and Tribulations' in Sharon Sandeen, Ansgar Ohly & Christopher Rademacher (eds) *Research Handbook on Information Law and Governance* (Elgar 2020)

Wiebe, A., 'Protection of industrial data – a new property right for the digital economy?' (2017) 12 JIPLP 62

Winkhoff, D., and Neumann, L.B., 'Towards a Social Contract for Genomics: Property and the Public in the 'Biotrust' Model' (2005) 1 Genomics, Society and Policy 8

Yew, R. J., 'Break It 'Til You Make It. An exploration o the Ramifications of Copyright Liability Under a Pre-training Paradigm of AI Development' (2024 Symposium on Computer Science and Law) 62

Yoo, C., 'Self-Actualization and the Need to Create as a Limit on Copyright' (2021) Faculty Scholarship at Penn Law 2041

Zeilinger, M., *Tactical Entanglements* (Meson Press 2021)

Ziaja, G., M., 'The Text and Data Mining Opt-out in Article 4(3) CDSMD: Adequate Veto Right for Rightsholders or a Suffocating Blanket for European Artificial Intelligence Innovations' (2024) JIPLP 1.