Governing AI for the Cultural Commons: Beyond Intellectual Property

Issue Brief

IT for Change

25 YEARS

unesco

In support of UNESCO

This research was developed under the AI, Culture and Intellectual Property Subgroup of the UNESCO Global Civil Society Organizations (CSO) and Academic Network on AI Ethics and Policy.

The Subgroup's activities have been developed within the framework of UNESCO's Global Civil Society Organisations (CSOs) and Academic Network on Artificial Intelligence (AI) Ethics and Policy, aiming to facilitate advocacy, knowledge exchange, and collaboration to widen civil society participation in the global AI governance landscape.

Recognizing the central role of knowledge and ideas in driving transformation, these activities reflect UNESCO's function as a laboratory of ideas. They have been developed and undertaken by independent CSOs and Academic Institutions and do not necessarily represent the views of UNESCO, its Member States, or its partners, nor do they commit the Organisation in any way.

Funding acknowledgement

This work was supported by the Patrick J. McGovern Foundation through the Scaling Inclusive, Human Rights-Based Ethical AI Governance (2025–26) project.

About IT for Change

Founded in 2000, IT for Change is a Southern NGO rooted in feminist principles and committed to advancing digital justice through the democratization of digital technologies. It holds Special Consultative Status with the United Nations Economic and Social Council (ECOSOC).

IT for Change

www.itforchange.net

In support of UNESCO

About the Global CSO and Academic Network on AI Ethics and Policy

Facilitated by UNESCO, the Network brings together CSOs and academic institutions from around the world to advance ethical, inclusive, and human rights–based approaches to AI governance. Launched in 2025 at the 3rd Global Forum on the Ethics of AI in Thailand, the Network serves as a platform for cooperation, knowledge exchange, and policy engagement, enabling its members to advance collaborative projects and partnerships to address the complex challenges posed by AI.

About the AI, Culture, and Intellectual Property Subgroup

The Subgroup examines the impact of AI on creative ecosystems, art production and related value-chains, as well as the preservation of digital and cultural commons. It is dedicated to formulating gender-responsive, rights-based policy recommendations to protect the digital cultural sovereignty of communities and creators, safeguard freedom of expression and artists' economic security, and to prevent the non-consensual commodification and militarisation of knowledge, including Indigenous knowledge forms.

Members of the AI, Culture, and Intellectual Property Subgroup

AI Impact Alliance (Co-Lead), Data Privacy Brazil, EiTESAL, European Law Institute, F-List for Music CIC, Homo Digitalis, InternetLab, IT for Change (Co-Lead), OdiseIA

Acknowledgements

Issue brief

Authors: Merrin Muhammed Ashraf and Shobhit S. (IT for Change)

Conceptualization, research framework, and review: Anita Gurumurthy and Nandini Chami (IT for Change)

Research support: Amoha Sharma (Center for Global Digital Justice) and Akshita Malviya (IT for Change)

Review and feedback: Achim Rosemann (UNESCO Ethics of AI Team) and Nina Houben and Klara Dreze (UNESCO Culture Sector)

Proofreading: Susan Sreemala (IT for Change)

Cover design and illustrations: Harikrishnan B. (IT for Change)

Report design: Niranjana Sivaram (IT for Change)

Design guidance: Chinmayi Arakali (IT for Change)

This issue brief is informed by the insights and perspectives emerging from the following:

Expert roundtable discussion

See the full list of participants in Annex 1.

Organizer: IT for Change

Facilitator: Nandini Chami (IT for Change)

Notetakers: Saloni Mishra and Merrin Muhammed Ashraf (IT for Change)

Regional focus group discussions

See the full list of participants in Annex 2.

Asia-Pacific

Organizer: IT for Change

Notetakers: Sadhana Sanjay, Saloni Mishra, and Akshita Malviya (IT for Change)

Africa

Organizer: IT for Change

Notetakers: Sadhana Sanjay, Saloni Mishra, and Akshita Malviya (IT for Change)

Ibero-American Region

Organizer: OdiseIA

Middle East and Africa

Organizer: EiTESAL

Think pieces

The following authors have contributed think pieces to the forthcoming series 'Knowledge Futures'.

Akshat Agrawal, AASA Chambers
Jai Vipra, Cornell University
Jim Thomas, ETC Group
Paul Keller, Open Future
Ramya Chandrasekhar, French National Centre for Scientific Research (CNRS)
Ranjit Singh, Data & Society
Rohini Lakshane, Independent technologist and researcher
Swaraj P. Barooah, SpicyIP
Victor Pavarin, Internet Lab

Contents

Executive Summary

Artificial Intelligence (AI) has emerged as a double-edged sword in the cultural space, especially after the advent of generative AI systems like ChatGPT. On one hand, AI, both generative and non-generative, presents opportunities for enhancing creative potential, lowering of entry barriers to produce art, and preserving cultural heritage. But on the other hand, the current AI innovation paradigm is driven by large-scale extraction and (mis)appropriation of creative and cultural resources into proprietary AI systems. This raises serious concerns surrounding not only the livelihoods of artists and creators, but also the collective cultural and developmental rights of communities and states, particularly in the Global South.

In response, current debates predominantly turn to prevailing Intellectual Property (IP) frameworks and their applicability to the AI moment. However, the new cultural production paradigm ushered by AI brings to sharp focus the limitations of the IP regime in safeguarding the legal and moral rights of creators. Moreover, it also points to how IP frameworks themselves contribute to the erosion of the knowledge commons by enabling large-scale extraction from the public domain and shared knowledge resources, while facilitating new forms of private enclosure over data and cultural materials.

This issue brief seeks to move beyond the largely celebratory narratives surrounding AI in the creative and cultural sectors.

This issue brief seeks to move beyond the largely celebratory narratives surrounding AI in the creative and cultural sectors. It aims to inform policymakers, regulators, and decision-makers across national and multilateral governance spaces—including culture, digital technology, IP, and trade—as well as civil society actors and cultural rights and development practitioners, about the emerging challenges posed by the dominant AI paradigm, by foregrounding the collective dimensions of the public domain and cultural data as critical developmental infrastructures. It examines how current AI development pathways are reshaping cultural production and knowledge systems, analyzes the role of IP frameworks in structuring these transformations, and explores emerging governance directions towards safeguarding epistemic justice and enabling fairer value sharing in AI-driven economies.

Section 1 introduces the issue brief, outlines its objectives and situates the broader debates around AI, cultural rights, and knowledge governance.

Section 2 details the methodology adopted to develop the brief through a structured and collaborative process among the members of the Subgroup on AI, Culture and Intellectual Property under the UNESCO Global Civil Society Organisation (CSO) and Academic Network on AI Ethics and Policy, combining desk research with stakeholder consultations.

Section 3 examines how the extractivist logic on which the dominant AI systems operate places pressure on the cultural commons and the conditions that sustain it. It explores how AI systems, trained on a vast corpus of creative and cultural material, threaten the displacement and devaluation of creative labor, pushing artists and creators to limit public sharing of their work, and potentially impoverishing the public domain. The section also examines how AI reshapes the representation and reproduction of cultural identity, with implications for the ability of communities to maintain and transmit their cultural heritage on their own terms. Further, it analyses how the current concentration of control over AI infrastructure and governance limits the ability of Global South countries and communities to shape technology trajectories in line with their developmental priorities.

Section 4 examines the role of IP frameworks in shaping this extractivist dynamic. It analyses how expansive interpretations of copyright exceptions enable data extraction at scale, while trade secrets, patents, and contractual arrangements are increasingly used to consolidate control over AI systems and their outputs.

Section 5 examines the limitations of conventional IP frameworks in responding to the new production paradigm of knowledge and cultural products that AI has ushered in. It argues that existing frameworks are ill equipped to safeguard creators' economic and moral rights in contexts where AI systems operate through aggregation and abstraction. The section further highlights how IP frameworks often overlook the incremental and collective nature of knowledge production, while lacking the mechanisms to sustain the broader informational and cultural ecosystems on which creative production depends.

Section 6 explores emerging directions for governance. It recognises that despite their conceptual limitations, IP frameworks must be reformed to prevent their use as vehicles for extraction and commodification. But more crucially, the section argues that broader governance approaches are needed to reclaim the knowledge infrastructures of tomorrow, by rejecting extractive innovation pathways towards approaches that are accountable to the communities whose knowledge systems they draw upon.

These include strengthening community data governance through stewardship-based approaches; evolving global frameworks for AI training data governance; building public AI infrastructures as a countervailing force to infrastructural concentration; and exploring redistributive mechanisms, including fiscal levies, to support the ecosystem and institutions that sustain cultural production and shared knowledge resources.

1. Introduction

Artificial intelligence (AI) systems are increasingly embedded in cultural production, knowledge creation, and information access. From creative tools to search interfaces and automated content generation, these systems are reshaping how knowledge is produced and circulated. In doing so, they are emerging not only as technological systems, but as knowledge infrastructures that influence whose knowledge is visible and who is able to participate in its production.

This transformation is closely tied to the large-scale use of cultural and informational resources in AI development. Training datasets draw extensively on publicly available content, open knowledge platforms, and collectively produced data, including traditional and Indigenous knowledge. While this has enabled rapid advances in AI capabilities, it has also raised questions about how such resources are accessed and governed. The large-scale incorporation of cultural materials into AI systems—often without meaningful mechanisms for participation or control—places pressure on the cultural commons and the conditions that sustain it.

The large-scale incorporation of cultural materials into AI systems—often without meaningful mechanisms for participation or control—places pressure on the cultural commons and the conditions that sustain it.

At the same time, AI governance is shaped by a dual dynamic. Expansive interpretations of copyright exceptions enable data extraction at scale, while proprietary controls—through trade secrets, patents, and contractual arrangements—are increasingly used to consolidate control over AI systems and their outputs. This dynamic shapes how value is generated and distributed, while also raising broader concerns around transparency, accountability, and ability to access or influence knowledge systems. That being said, it must be also acknowledged that the exceptions in IP frameworks—such as copyright and patent—have been crucial for the growth of startups and technology ecosystems including in developing/middle income countries.

These developments intersect with growing concerns around digital cultural sovereignty. As AI systems become central to knowledge production, the ability of communities and creators to maintain control over their cultural resources becomes more significant. Parallelly, efforts to restrict data access in response to AI scraping are creating new forms of enclosure, potentially limiting the growth of open and shared knowledge systems and reinforcing existing asymmetries in technological capacity.

The implications of these shifts are particularly significant from a development perspective. AI systems are reshaping creative labor markets and influencing the visibility and representation of different cultures and knowledge systems. These processes are occurring alongside a growing concentration of power in the infrastructures that organize knowledge production and access, raising questions about participation and the distribution of benefits in the AI economy.

Yet these trajectories are not inevitable. AI systems also hold the potential to support more inclusive and diverse knowledge ecosystems when developed and governed in ways that strengthen the commons. Across contexts, AI tools are being explored for applications such as language revitalization, cultural archiving, translation, and accessibility, including for historically underrepresented linguistic and cultural communities. These developments point to the possibility of AI innovation that broadens participation in cultural and knowledge production, where such innovation is supported by appropriate governance frameworks. The challenge, therefore, is not AI use in the cultural domain *per se*, but the conditions under which AI systems are developed, integrated into cultural economies and governed.

This issue brief examines these dynamics at the intersection of AI, IP, and cultural rights. It analyses how current AI development pathways are reshaping cultural production and knowledge systems, the role of IP frameworks in enabling and structuring these processes, and the implications for the cultural commons and public domain. The brief's critique of IP frameworks is not to discount their role in supporting innovation and industrial development in the Global South. Rather, it is making a case against the uncritical application of conventional IP frameworks in the new production paradigm of knowledge and cultural products ushered in and shaped by an AI economy that is concentrated in the hands of a few dominant actors. In this context, the brief explores emerging directions for governance, including approaches to data stewardship, public AI infrastructure, and alternative mechanisms for redistributing value.

By situating AI within broader questions of knowledge governance and development, the brief aims to contribute to ongoing discussions on how AI systems can be shaped in ways that sustain diverse knowledge systems, support collective cultural rights, and enable more equitable participation of communities and Global South states in the development and use of AI.

2. Methodology

This issue brief was developed collaboratively by the members of the Subgroup on AI, Culture and Intellectual Property under the UNESCO Global Civil Society Organisation (CSO) and Academic Network on AI Ethics and Policy. The writing of the issue brief was informed by a combination of desk research and stakeholder consultations through an expert roundtable and regional focus group discussions. The consultations with stakeholders and experts probed three major aspects:

1. The impacts of the dominant AI innovation paradigm on cultural rights and development
2. The adequacy of existing IP regimes in shaping AI development pathways towards local needs and control over knowledge infrastructures
3. The role of the AI commons and IP reform in sustaining a flourishing public domain.

Each of the activities undertaken towards the development of the brief are explained in this section.

Expert roundtable discussion

The roundtable was organized on the sidelines of the India AI Impact Summit in New Delhi in February 2026. It brought together policy and legal experts, creators and artists, researchers, and representatives from digital rights and development organizations from different regions of the world to examine the public domain challenge within ongoing policy debates on AI innovation, cultural rights, and development.[1] The discussion was structured around three rounds of catalyst presentations followed by open exchange among participants.

Think pieces from experts

We invited nine roundtable participants to write short, provocative think pieces that explore the different dimensions of the cultural-rights impact of the current extractivist AI paradigm, with the aim of bringing in diverse expert perspectives on the issues under consideration in this brief. The think pieces have been accepted for publication by the magazine Bot Populi as a special series titled 'Knowledge Futures'.[2]

1 See Annex 1 for the list of participants of the roundtable.

2 Bot Populi. (2026, forthcoming). Knowledge Futures. Bot Populi. https://botpopuli.net/section/knowledge-futures/

Regional focus group discussions

To further examine these issues across different regional contexts, four virtual Focus Group Discussions (FGDs) were conducted covering the Asia-Pacific, Africa (two FGDs), and the Ibero-American regions.[3] In addition to probing the aspects mentioned above, these discussions explored, in particular, forms of resistance and responses by communities seeking to assert control over their knowledge systems and cultural expressions.

Literature review

An initial literature review was undertaken to map the evolving landscape of debates on the issue, identify key conceptual and policy tensions, and develop an analytical framework to guide the research. Subsequently, the inputs from the roundtable, FGDs and reflections from the think pieces were consolidated and synthesized to identify key themes, areas of convergence, and points of divergence. These findings were supplemented through a targeted review of relevant academic literature and policy documents.

3 See Annex 2 for the list of participants of each regional focus group discussion.

3. AI Extractivism and its Implications for Cultural Production

Section summary

The current model of AI innovation is reconfiguring cultural production and knowledge systems by extracting from the commons, restructuring labor, and concentrating control over knowledge infrastructures.

- The knowledge commons are being reshaped into a continuous input layer for AI model development, rather than a bounded system of shared production. This is placing pressure on the economic incentives and institutional conditions under which the commons are produced and maintained.
- As AI systems draw on large bodies of existing creative work while generating outputs that can substitute commissioned or licensed work, creative labor is being restructured, with reduced demand and lower value for commissioned work.
- AI is reconstituting cultural knowledge by privileging standardization and generalization, leading to the homogenization of cultural expression and the marginalization of plural knowledge systems.
- While cultural resources, labor, and knowledge from across the world are incorporated into AI systems, decision-making power over their design, deployment, and governance remains concentrated in the hands of the few.
- These factors limit the ability of countries and communities—particularly in the Global South—to shape technological trajectories or ensure that these systems reflect their values and context, undermining their right to development.

The integration of AI systems into cultural and knowledge processes is already producing observable shifts across multiple domains. These shifts extend to affect how creative work is organized, how knowledge is represented, how value is generated, and how participation in cultural production is structured.

This section examines these developments across four interrelated dimensions: (i) the extraction and enclosure of the cultural commons; the restructuring of creative labor and livelihoods; (ii) the disembedding and homogenization of cultural knowledge; (iii) the implications for the cultural commons and public domain; and (iv) the developmental concerns arising from the concentration of epistemic power. Taken together, these dynamics illustrate how current AI development pathways are reshaping the conditions under which knowledge is produced, shared, and sustained.

3.1. Extraction and enclosure of the cultural commons

AI systems draw on large training datasets assembled from publicly available cultural material at scale. Contemporary large language models (LLMs) are known to absorb data from open source repositories such as Common Crawl (estimated to contain billions of web pages), Wikipedia, digitized books, open-access journals, and large code repositories like GitHub.[4] These sources reflect different modes of collective production—volunteer-driven knowledge infrastructures, publicly funded research outputs, and open licensing ecosystems—each with their own norms surrounding attribution, reuse, and public benefit. As participants highlighted in the Africa FGDs and the Asia-Pacific FGD, the incorporation of these materials into AI systems proceeds through processes that operate independently of the institutional frameworks governing their production.

In the case of Wikipedia, content produced through ongoing editorial labor and community oversight is ingested into AI training pipelines without mechanisms for attribution or feedback. Wikimedia Foundation has raised concerns about sustained, automated scraping linked to AI development, resulting in appropriation without any reciprocal value flows.[5] Such peer-produced repositories have become foundational to model training precisely because they offer high-quality, structured, and continuously updated text at scale.[6]

4 Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C.,...,Amodei, D. (2020). Language models are few-shot learners. arXiv. https://doi.org/10.48550/arXiv.2005.14165; Lakshane, R. (2026, forthcoming). LLM Extractivism and the Politics of the Knowledge Commons. In Knowledge Futures. Bot Populi https://botpopuli.net/section/knowledge-futures/

5 Ndubane, D. (2026, focus group discussion). Oral inputs at the regional focus group discussion-Middle East & Africa: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 11 2026).

6 Lakshane, R. (2026, forthcoming). LLM Extractivism and the Politics of the Knowledge Commons. In Knowledge Futures. Bot Populi https://botpopuli.net/section/knowledge-futures/; Lakshane, R. (2026, February 17). Challenges for cultural rights and developmentin the dominant AI innovation Paradigm. In Cultural Rights,Innovation, and Development in the AI moment: Towards a public domain framing. IT for Change. https://itforchange.net/event-report-cultural-rights-innovation-and-development-ai-moment-towards-a-public-domain-framing/

Beyond these digital commons, similar dynamics extend to digitized forms of intangible cultural heritage. Linguistic datasets, oral traditions, community archives and cultural repositories—often created through publicly funded research or community documentation initatives—are increasingly incorporated into training datasets.[7] Such materials, often produced within specific cultural and institutional contexts, are absorbed into AI systems without mechanisms for community consent, attribution, control over downstream use or benefit sharing.[8] In this process, knowledge that was historically embedded in particular social and cultural settings is recast as general-purpose training data.

In this process, knowledge that was historically embedded in particular social and cultural settings is recast as general-purpose training data.

Similar tensions arise due to the misalignment between commons-based licensing regimes and large-scale model training processes. Licenses such as Creative Commons are designed to enable reuse under specific conditions relating to attribution, non-commercial use and share-alike requirements. However, as Chandrasekhar notes, once content under such licenses is incorporated into model parameters, the technical architecture of AI systems does not support the tracing or enforcement of these conditions.[9] This effectively detaches reuse from the terms under which the knowledge is shared.

In software ecosystems too, large datasets used to train code-generating models draw extensively on repositories governed by open-source software licenses such as General Public License and MIT License. These licenses impose obligations relating to attribution and reuse, but these obligations become almost impossible to operationalize once code is absorbed into training corpora and reproduced in generated outputs. Ongoing litigation, including cases against GitHub Copilot, centres on precisely these questions of attribution, licensing compliance, and the boundary between reuse and reproduction in AI systems.[10]

---

7 Cherukuri, K. S., Moses, P. A., Sakata, A., Chen, J., & Chen, H. (2025). Large language models for oral history understanding with text classification and sentiment analysis. arXiv. https://doi.org/10.48550/arXiv.2508.06729 ; Participant 1. (2026, focus group discussion). Oral inputs at the regional focus group discussion-Middle East & Africa: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 11 2026)

8 UNESCO. (2025, November 28). A new expert report explores how AI is transforming culture. UNESCO. https://www.unesco.org/en/articles/new-expert-report-explores-how-ai-transforming-culture; Vogliano, S. (2026, focus group discussion). Oral inputs at the regional focus group discussion- Ibero-America: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 16 2026), which points to the increasing appropriation of biodiversity data, genetic information and traditional knowledge for training AI systems. This process often occurs without meaningful participation or benefit-sharing, particularly for Indigenous communities, representing a form of biopiracy.

9 Chandrasekhar, R. (2026, forthcoming). Nurturing AI commons: Reflections from emerging licensing initiatives for training datasets. In Knowledge Futures. Bot Populi https://botpopuli.net/section/knowledge-futures/; Chandrasekhar, R. (2026, February 17). Legal-policy pathways to enrich the data and AI commons. In Cultural Rights,Innovation, and Development in the AI moment: Towards a public domain framing. IT for Change. https://itforchange.net/event-report-cultural-rights-innovation-and-development-ai-moment-towards-a-public-domain-framing/

10 Joseph Saveri Law Firm & Butterick, M. (2022, November 3). GitHub Copilot litigation. https://githubcopilotlitigation.com/

The cumulative result of this AI-driven appropriation is the reconfiguration of the commons as a continuous input layer for model development, rather than a bounded system of shared production. This shift is already placing pressure on the infrastructures that sustain the commons. Wikimedia, for instance, has reported an increasing server load due to automated scraping;[11] and news organizations have pointed to declining traffic linked to the easy availability of AI-generated summaries.[12]

The cumulative result of this AI-driven appropriation is the reconfiguration of the commons as a continuous input layer for model development, rather than a bounded system of shared production.

Over time, these dynamics place pressure on the conditions under which the commons is produced and maintained. As the relationship between contribution and reuse becomes increasingly decoupled, the institutional arrangements that sustain collective knowledge production come under strain, raising questions about their long-term viability.

3.2. Restructuring of creative labor and livelihoods

As AI systems are integrated into creative production and distribution, artists and creators are encountering changes in how their work is valued, remunerated, and circulated. AI systems draw on large bodies of existing creative work while generating outputs can substitute commissioned or licensed work, affecting both demand and pricing structures in downstream markets.[13]

In visual art and design, illustrators have reported a reduction in commissioned work as clients increasingly rely on AI-generated images for concept development, prototyping, and, in some cases, final outputs.[14]

---

11 Miller, M. (2025, October 17). New user trends on Wikipedia. Wikimedia Foundation. https://diff.wikimedia.org/2025/10/17/new-user-trends-on-wikipedia/; Mueller, B., Danis, C., & Lavagetto, G. (2025, April 1). How crawlers impact the operations of the Wikimedia projects. Wikimedia Foundation. https://diff.wikimedia.org/2025/04/01/how-crawlers-impact-the-operations-of-the-wikimedia-projects/; Khosravi, M., & Yoganarasimhan, H. (2026). Impact of AI search summaries on website traffic: Evidence from Google AI overviews and Wikipedia. arXiv. https://doi.org/10.48550/arXiv.2602.18455

12 Savage, M. (2025, July 24). AI summaries causing 'devastating' drop in online news audiences, study finds. The Guardian. https://www.theguardian.com/technology/2025/jul/24/ai-summaries-causing-devastating-drop-in-online-news-audiences-study-finds

13 UNESCO. (2026, February 18). Creators face projected global revenue losses of up to 24% by 2028, new UNESCO report shows. UNESCO. https://www.unesco.org/en/articles/creators-face-projected-global-revenue-losses-24-2028-new-unesco-report-shows

14 O'Connor, S., & Burn-Murdoch, J. (2026, February 12). The AI shift: What has AI done to illustrators? Financial Times. https://www.ft.com/content/98c48b7c-f07c-4be2-a769-715582fa09ca; The Society of Authors. (2024, April 11). SoA survey reveals a third of translators and quarter of illustrators losing work to AI. Society of Authors. https://societyofauthors.org/2024/04/11/soa-survey-reveals-a-third-of-translators-and-quarter-of-illustrators-losing-work-to-ai/

This is particularly visible in commercial illustration and design workflows, where AI rolls are used to produce multiple iterations at low cost, reducing both the volume of assignments and the rates offered. Rather than eliminating labor entirely, this shifts artists toward smaller, more fragmented tasks, often centered on refining or editing machine-generated outputs. Legal actions brought by artists against companies such as Stability AI and Midjourney, as well as by Getty Images over the use of its photo archive, reflect these concerns about how artistic work is assimilated into training datasets and then re-enters the market in altered form.[15]

Similar pressures are visible in writing and journalism. Independent writers and media workers report that clients and platforms are increasingly using AI systems to generate drafts, summaries, and routine content, with human labor repositioned towards verification and editing.[16] These changes are affecting commissioning practices, with fewer assignments for original writing and tighter timelines for delivery. Emerging research indicates disproportionate impact on women, with a recent survey suggesting that women creative workers are losing revenue at nearly twice the rate of men.[17] The consequences are also particularly dire for independent and regional media outlets, where declining revenues already constrain the ability to sustain original reporting.

These developments are prompting new forms of worker mobilization across cultural sectors. Writers, actors, visual artists, musicians, and other media workers have started to organize around demands relating to consent, attribution, compensation, and protections against AI-driven devaluation and displacement.[18] The spate of legal actions, strikes, collective bargaining efforts, and media campaigns across geographies reflects growing concerns surrounding the incorporation of creative work into AI production systems.

The spate of legal actions, strikes, collective bargaining efforts, and media campaigns across geographies reflects growing concerns surrounding the incorporation of creative work into AI production systems.

15 Getty Images (US) Inc & Ors v Stability AI Ltd [2025] EWHC 2863 (Ch) (High Court of England and Wales, Chancery Division, November 4, 2025). https://www.bailii.org/ew/cases/EWHC/Ch/2025/2863.pdf; Schor, Z. (2024, December 2). Andersen v. Stability AI: The landmark case unpacking the copyright risks of AI image generators. NYU Journal of Intellectual Property & Entertainment Law. https://jipel.law.nyu.edu/andersen-v-stability-ai-the-landmark-case-unpacking-the-copyright-risks-of-ai-image-generators/

16 Portelli, A. (2026, focus group discussion). Oral inputs at the regional focus group discussion- Asia-Pacific: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 10 2026).; Thurman, N., Thäsler-Kordonouri, S., & Fletcher, R. (2025, November 27). AI adoption by UK journalists and their newsrooms: Surveying applications, approaches, and attitudes. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/ai-adoption-uk-journalists-and-their-newsrooms-surveying-applications-approaches-and-attitudes

17 Institute for the Future of Work (2025). Creative Industries and GenAI: Good Work impacts on a sector in transition. London: Institute for the Future of Work. DOI: 10.5281/zenodo.15535400

18 ibid

Moreover, the incorporation of creative work into AI training pipelines is beginning to affect how, and to what extent, creative work is shared. Artists and creators are becoming more cautious about making their work publicly available, particularly in open or discoverable formats, given the likelihood that it may be incorporated into training datasets.[19] Some have adopted technical measures such as watermarking or started using access-restriction tools,[20] while also considering withdrawing their content from platforms altogether.

The effects of these changes extend beyond individual livelihoods to the broader ecosystem of cultural production. As AI systems draw on existing creative work while simultaneously competing with it, they weaken the economic incentives and institutional conditions that sustain ongoing production. This also feeds back into the broader cultural ecosystem as creators start withdrawing or restricting their work from the public domain, the volume and diversity of the public domain is reduced, with grave implications for future cultural production.

As AI systems draw on existing creative work while simultaneously competing with it, they weaken the economic incentives and institutional conditions that sustain ongoing production.

3.3. Disembedding and homogenization of cultural knowledge

As cultural resources are integrated into AI systems, they are separated from the social, political, and historical contexts in which they were produced. This process of disembedding is particularly significant for forms of knowledge that are embedded in community practices and cultural norms—especially those of Indigenous peoples and other marginalized groups.[21]

---

19 Yahaya, M.S. (2026, focus group discussion). Oral inputs at the regional focus group discussion-Middle East & Africa: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 11 2026).

20 Narayan, V. (2026, focus group discussion). Oral inputs at the regional focus group discussion- Asia-Pacific: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 10 2026); University of Chicago. (n.d.). What is Nightshade? https://nightshade.cs.uchicago.edu/whatis.html; University of Chicago. (n.d.). What is Glaze? https://glaze.cs.uchicago.edu/what-is-glaze.html

21 UNESCO. (2025, November 25). A new expert report explores how AI is transforming culture. UNESCO. https://www.unesco.org/en/articles/new-expert-report-explores-how-ai-transforming-culture; Kasosi, L. (2025, August 8). Indigenous peoples and AI: Defending rights, shaping future technology. Cultural Survival. https://www.culturalsurvival.org/news/indigenous-peoples-and-ai-defending-rights-shaping-future-technology

Linguistic datasets drawn from African and South Asian languages are increasingly incorporated into AI systems as standardized inputs, even where those languages are tied to specific oral traditions, social relations, and contextual modes of meaning-making.[22] The semantic reworking of *Batik* prints from Indonesia offer another case in point, where culturally specific motifs are reproduced in AI-generated outputs without reference to their cultural significance.[23] In such cases, cultural forms are treated as aesthetic patterns rather than carriers of meaning, resulting in what has been described as a form of "semantic dispossession"—the loss of communities' control over how their cultural expressions are interpreted and circulated.[24]

[C]ultural forms are treated as aesthetic patterns rather than carriers of meaning, resulting in what has been described as a form of "semantic dispossession"—the loss of communities' control over how their cultural expressions are interpreted and circulated.

The issue extends beyond discrete instances of misrepresentation to implicate how cultural visibility itself is mediated. When recognition and circulation become tied to algorithmic systems, cultural forms are filtered through criteria of legibility—what can be processed, standardized, and reproduced within model architectures. As Ranjit Singh argues, the result is the production of a "baseline" version of culture, where dominant languages and styles are more readily encoded and reproduced, while others are more likely to be flattened, misrepresented, or excluded.[25]

As AI systems increasingly mediate access to knowledge, they begin to reinforce these hierarchies through iterative cycles of retrieval and recombination. What is already well-represented becomes easier to retrieve, and what is easier to retrieve becomes more likely to be reproduced. Andrew Peterson characterizes this as a form of "knowledge collapse," where the range of accessible knowledge narrows over time as systems preferentially surface content that is already legible to them.[26]

---

22 Sokatsha, Z. (2026, focus group discussion). Oral inputs at the regional focus group discussion-Middle East & Africa: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 11 2026); Deborah,I.C. (2026, focus group discussion). Oral inputs at the regional focus group discussion- Asia-Pacific: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 10 2026); Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21) (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

23 Deborah,I.C. (2026, focus group discussion). Oral inputs at the regional focus group discussion- Asia-Pacific: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 10 2026)

24 ibid.

25 Singh, R. (2026, forthcoming). Enclosure of Legibility. In Knowledge Futures. Bot Populi https://botpopuli.net/section/knowledge-futures/; Singh, R. (2026, February 17). Challenges for cultural rights and developmentin the dominant AI innovation Paradigm. In Cultural Rights,Innovation, and Development in the AI moment: Towards a public domain framing. IT for Change. https://itforchange.net/event-report-cultural-rights-innovation-and-development-ai-moment-towards-a-public-domain-framing/

26 Peterson, A. J. (2024). AI and the problem of knowledge collapse. arXiv. https://doi.org/10.48550/arXiv.2404.03502

Cultural forms and knowledge systems that are less digitized, less standardized, or embedded in oral and community-based practices are progressively rendered less visible within these systems, because they fall outside the conditions of "algorithmic legibility."[27] These concerns are also echoed in the UNESCO Recommendation on the Ethics of AI, which warns that natural language processing (NLP) systems could contribute to "the disappearance of endangered languages, local dialects, and tonal and cultural variations associated with human language and expression."[28]

This is reflected in model performance disparities across languages and dialects: widely digitized languages receive more accurate outputs, while others are translated inconsistently, simplified, or excluded.[29] At the same time, when marginal cultural expressions are assimilated, they are often reproduced in generic ways that strip them from their social and historical grounding.[30] This produces a condition in which some cultural forms are rendered invisible, while others are reproduced only after being stripped of their contextual meaning, making it difficult for affected communities to detect or contest how their cultures are rendered within AI systems.

The result is a gradual reconstitution of cultural knowledge within AI systems that privileges standardization and generalization. Over time, this contributes to the homogenization of cultural expression and the marginalization of plural knowledge systems.[31] It also reshapes how cultural identity is represented and reproduced, with implications for the ability of communities to maintain and transmit their cultural heritage on their own terms.

---

27 Singh, R. (2026, forthcoming). Enclosure of Legibility. In Knowledge Futures. Bot Populi https://botpopuli.net/section/knowledge-futures/; Singh, R. (2026, February 17). Challenges for cultural rights and developmentin the dominant AI innovation Paradigm. In Cultural Rights,Innovation, and Development in the AI moment: Towards a public domain framing. IT for Change. https://itforchange.net/event-report-cultural-rights-innovation-and-development-ai-moment-towards-a-public-domain-framing/; Lakshmi, D. S. (2025). Rebranding empire in the age of generative AI. Frontiers in Communication. https://doi.org/10.3389/fcomm.2025.1604361

28 UNESCO. (2023, May 16). Recommendation on the ethics of artificial intelligence. UNESCO (Policy Area 7, para. 95). https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

29 Li, Z., Shi, Y., Liu, Z., Yang, F., Payani, A., Liu, N., & Du, M. (2024). Quantifying multilingual performance of large language models across languages. arXiv. https://doi.org/10.48550/arXiv.2404.11553

30 McMullan, J., & Stasiuk, G. (2025, March 12). How AI images are "flattening" Indigenous cultures, creating a new form of tech colonialism. The Conversation. https://theconversation.com/how-ai-images-are-flattening-indigenous-cultures-creating-a-new-form-of-tech-colonialism-246972; Turk, V. (2023, October 10). How AI reduces the world to stereotypes. Rest of World. https://restofworld.org/2023/ai-image-stereotypes/; Zhao, Z., & Song, J. (2026). Reimagining mythological landscapes: Leveraging AIGC to visualize the Classic of Mountains and Seas (Shan Hai Jing). Digital Scholarship in the Humanities. https://doi.org/10.1093/llc/fqag0177

31 Singh, R. (2026, forthcoming). Enclosure of Legibility. In Knowledge Futures. Bot Populi https://botpopuli.net/section/knowledge-futures/; Singh, R. (2026, February 17). Challenges for cultural rights and developmentin the dominant AI innovation Paradigm. In Cultural Rights, Innovation, and Development in the AI moment: Towards a public domain framing. IT for Change. https://itforchange.net/event-report-cultural-rights-innovation-and-development-ai-moment-towards-a-public-domain-framing/ ;UNESCO. (2025, November 25). A new expert report explores how AI is transforming culture. UNESCO. https://www.unesco.org/en/articles/new-expert-report-explores-how-ai-transforming-culture;Organisation for Economic Co-operation and Development (OECD). (2024). Towards substantive equality in artificial intelligence: Transformative AI policy for gender equality and diversity. OECD AI. https://wp.oecd.ai/app/uploads/2025/05/towards-substantive-equality-in-artificial-intelligence_Transformative-AI-policy-for-gender-equality-and-diversity.pdf; Office of the United Nations High Commissioner for Human Rights. (2025, July 30). Artificial intelligence and creativity (A/80/278) [Report of the Special Rapporteur in the field of cultural rights, Alexandra Xanthaki]. OHCHR. https://www.ohchr.org/en/documents/thematic-reports/a80278-artificial-intelligence-and-creativity-report-special-rapporteur

3.4. Concentration of epistemic power and development constraints

The processes described above are firmly embedded within a broader reorganization of knowledge and power. AI systems function as infrastructures through which knowledge is produced, accessed, and circulated over time. They shape how learning takes place, which forms of knowledge are prioritized, and how capabilities are built across different parts of the global economy. As Laura Mann notes, such systems increasingly operate as infrastructures that structure learning and innovation, rather than merely intermediating information flows.[32]

In the current AI paradigm, control over these infrastructures is concentrated among a small number of firms and institutions with access to large-scale data, compute, and technical expertise.[33] A very small set of firms currently develop and deploy frontier AI models, supported by access to proprietary datasets and high-performance computing infrastructure. This concentration allows these actors to determine how knowledge is captured, organized, and deployed, while also shaping who is able to learn from these systems. Where access to models, data and computational resources is restricted, the ability of actors—particularly from the Global South—to build technical capabilities and develop alternative systems is correspondingly limited.[34] As a result, AI systems not only extract value from distributed knowledge resources, but also weaken the conditions for local learning, innovation, and capital accumulation.[35]

AI systems developed within a narrow set of institutional and geographic contexts embed particular assumptions about language, value, and relevance into their design.[36] As these systems are deployed globally, these assumptions travel with them, influencing what counts as knowledge, which forms of expertise are recognized, and how cultural and economic activity is organized.

---

32 Mann, L., & Iazzolino, G. (2019). Digital platforms as privatized epistemic infrastructures. IT for Change. https://projects.itforchange.net/platformpolitics/wp-content/uploads/2019/03/Digital-Platforms-as-Privatized-Epistemic-Infrastructures-_5thMarch.pdf

33 United Nations Conference on Trade and Development. (2025). Technology and innovation report 2025: Inclusive artificial intelligence for development.UNCTAD. https://unctad.org/publication/technology-and-innovation-report-2025; United Nations Conference on Trade and Development. (2024). Digital economy report 2024: Shaping an environmentally sustainable and inclusive digital future. UNCTAD. https://unctad.org/publication/digital-economy-report-2024

34 González, A.P. (2026, focus group discussion). Oral inputs at the regional focus group discussion- Ibero-America: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 16 2026)

35 IT for Change, & Center for Global Digital Justice. (2025, December). Joint submission to the call for input for the EMRTD study “Artificial intelligence, cultural rights, and the right to development”. IT for Change. https://itforchange.net/joint-submission-to-call-for-input-for-emrtd-study-%e2%80%98artificial-intelligence-cultural-rights-and/ ; Alberto. (2026, focus group discussion). Oral inputs at the regional focus group discussion- Ibero-America: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 16 2026)

36 Pavarin, V. (2026, forthcoming). Cultural Sovereignty in the AI age. In King” (New Delhi). In Event report: IP and culture roundtable. IT for Change. Knowledge Futures. Bot Populi https://botpopuli.net/section/knowledge-futures/; Pavarin, V. (2026, February 17). Challenges for cultural rights and developmentin the dominant AI innovation Paradigm. In Cultural Rights, Innovation, and Development in the AI moment: Towards a public domain framing. IT for Change. https://itforchange.net/event-report-cultural-rights-innovation-and-development-ai-moment-towards-a-public-domain-framing/

In this sense, AI systems contribute to a reordering of knowledge relations, in which control over the infrastructures of learning and innovation remains concentrated even as resources are drawn from a wide range of contexts. Michael Kwet describes this as a form of "digital colonialism," where data and knowledge resources from the Global South are extracted and integrated into systems controlled in the North, reinforcing technological dependence and asymmetries in the global knowledge economy.[37]

AI systems contribute to a reordering of knowledge relations, in which control over the infrastructures of learning and innovation remains concentrated even as resources are drawn from a wide range of contexts.

The concentration of control over these infrastructures directly engages questions of cultural rights. The Committee on Economic, Social and Cultural Rights (CESCR) has affirmed the right of peoples to enjoy the benefits of scientific progress, and its applications, including through participation in decision making relating to science and technology.[38] It further emphasises that scientific progress must be directed toward the "needs of the population" and be accessible and relevant, particularly for disadvantaged groups.[39] This establishes a clear normative expectation that scientific and technological development be shaped through inclusive governance and be aligned with social priorities.

The current organization of AI development constrains the exercise of these rights. Cultural resources, labor, and knowledge from across the world are incorporated into AI systems, but decision-making power over their design, deployment, and governance remains concentrated. Participants across African, Middle Eastern, and Ibero-American discussions noted that this limits the ability of countries and communities—particularly in the Global South—to participate meaningfully in shaping technological trajectories or ensure that these systems reflect their needs, values, and contexts. In effect, the same systems that draw on situated knowledge and cultural practices also displace or devalue them—for instance, by substituting local, experience-based knowledge systems with standardized, model-driven outputs—further constraining opportunities for endogenous learning and innovation.[40]

---

37 Kwet, M. (2018, August 15). Digital colonialism: US empire and the new imperialism in the Global South. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3232297; Botero, C. (2026, focus group discussion). Oral inputs at the regional focus group discussion- Ibero-America: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 16 2026)

38 United Nations Committee on Economic, Social and Cultural Rights. (2020). General comment No. 25 (2020) on science and economic, social and cultural rights (article 15 (1) (b), (2), (3) and (4) of the International Covenant on Economic, Social and Cultural Rights). UNCESCR. https://docs.un.org/en/e/c.12/gc/25

39 United Nations Committee on Economic, Social and Cultural Rights. (2020, April 30). General comment No. 25 (2020) on science and economic, social and cultural rights (article 15 (1) (b), (2), (3) and (4) of the International Covenant on Economic, Social and Cultural Rights).UNCESCR. https://docs.un.org/en/e/c.12/gc/25

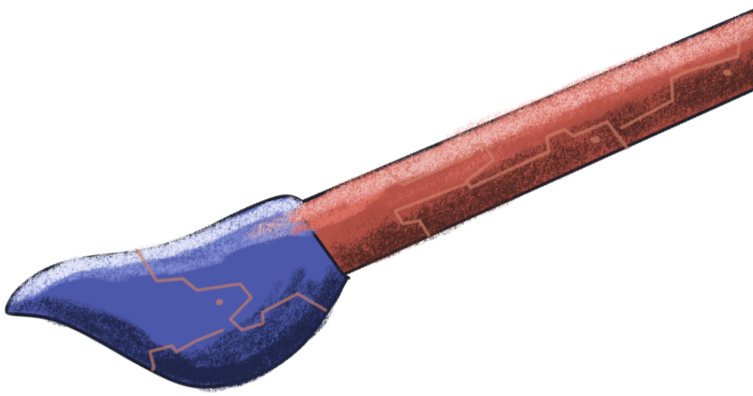
40 IT for Change, & Center for Global Digital Justice. (2025, December). Joint submission to the call for input for the EMRTD study "Artificial intelligence, cultural rights, and the right to development". IT for Change. https://itforchange.net/joint-submission-to-call-for-input-for-emrtd-study-%e2%80%98artificial-intelligence-cultural-rights-and/

This constitutes a clear challenge to the right to development. The UN Declaration on the Right to Development affirms that “all peoples are entitled to participate in, contribute to, and benefit from economic, social, and cultural development,” and that development must ensure the “active, free and meaningful participation” of people and the “fair distribution of the benefits resulting therefrom.”[41]

Similarly, the 2005 Convention on the Protection and Promotion of the Diversity of Cultural Expressions recognizes culture as “one of the mainsprings of development.”[42] The current configuration of the AI economy undermines these conditions.

Where participation is reduced to data provision, contribution is decoupled from control, and benefits accrue elsewhere, the conditions required to realize the right to development are not met.

As shown in this section, the appropriation of the commons, displacement of creative labor, and the reconstitution of cultural knowledge, cumulatively reconfigure how value is generated and distributed. Where participation is reduced to data provision, contribution is decoupled from control, and benefits accrue elsewhere, the conditions required to realize the right to development are not met.



41 United Nations General Assembly. (1986, December 4). Declaration on the right to development (A/RES/41/128). OHCHR.https://www.ohchr.org/en/instruments-mechanisms/instruments/declaration-right-development; United Nations General Assembly. (1986, December 4). Declaration on the right to development (A/RES/41/128). OHCHR. https://legal.un.org/avl/pdf/ha/drd/drd_e.pdf

42 United Nations Educational, Scientific and Cultural Organization. (2005, 20 October). Convention on the protection and promotion of the diversity of cultural expressions. Article 2, Principle 5

4. Intellectual Property and the Dynamics of AI Extractivism

Section summary

Current IP regimes are central to enabling AI extractivism through a twofold process, where the extraction of copyrighted materials, shared knowledge resource, and cultural commons is followed by its enclosure via proprietary models and the abuse of trade secrets and patents.

- Copyright exceptions ("fair use," "fair dealing" and "text and data mining" (TDM)), are enabling the large-scale extraction of protected works without consent, authorization and reciprocity.
- The fairness of such claims is highly contested, particularly given the commercial orientation of most AI systems and their potential to displace creative labor by generating competing outputs at a previously unimaginable speed and scale.
- Patent and trade secrets are mobilized by dominant AI actors to assert proprietary claims and consolidate control over knowledge products derived from both protected works and the cultural commons.
- These dynamics are enabling a concentration of control over knowledge systems built on widely shared cultural resources and, more broadly, over the knowledge economy.

In this extractive AI paradigm, IP regimes, particularly copyright law, have come into sharp focus for their inability to safeguard the legal and moral rights of people over their creative works. With increasing reliance of AI developers on copyrighted material for training leading AI models, their claim to fair use exception under copyright law is being fiercely contested by copyright holders in courts across jurisdictions. This is only one side of the story. IP regimes not only aid in the expropriation of creative labor, but are also being mobilized by powerful AI actors to claim proprietary rights and monopoly over knowledge products derived from such expropriated work, as well as from the extraction of the knowledge commons and collective cultural production. This propertization of innovation goods coupled with the impoverishment of the commons as discussed in the previous section creates what Akshat Agrawal terms as "double enclosure."[43]

43 Agrawal, A. (2026, forthcoming). Beyond Copyright: Why IP regimes cannot address AI's displacement of creative labor. In Knowledge Futures. Bot Populi https://botpopuli.net/section/knowledge-futures/; Agrawal, A. (2026, February 17). IP Regimes, mainstream AI Innovation models, and implications for the public domain. In Cultural Rights, Innovation, and Development in the AI moment: Towards a public domain framing. IT for Change. https://itforchange.net/event-report-cultural-rights-innovation-and-development-ai-moment-towards-a-public-domain-framing/

In this section, we briefly examine how the current copyright law, trade secrets, and patent regimes enable powerful AI actors to entrench extractivism, market concentration, and enclosure of innovations.

4.1. Copyright exceptions and the expansion of data extraction

A UK government report noted that "AI developers have stated that it would be impossible to train today's leading AI models without using copyrighted materials."[44] This includes copyrighted material obtained through various licensing arrangements as well as those scraped from the open web. In case of the latter, creative works are funneled into AI training circuits without the knowledge or authorization of copyright holders. The lack of transparency of AI training datasets further thwarts the ability of creators to identify whether their work has been used and claim fair compensation for the reuse of their work[45]. AI developers, on their part, rely on exemptions within the copyright regime that many jurisdictions provide—such as Text and Data Mining (TDM) exceptions subject to different conditionalities or restrictions, or a broader fair use[46] or fair dealing[47] exception that can be adapted to TDM research—to build training datasets.[48] While claims to these exceptions *vis-a-vis* rights copyright holders have always been contentious, *en masse* TDM mining by Big Tech and resource-rich AI actors has alarmed content creators like never before, leading to a flurry of lawsuits against AI companies, challenging their claims to copyright exceptions.

AI-related fair use industries alone generated $1.7 trillion out of $10.2 trillion revenue generated by fair use industries in the US, a rise of 78% since 2017.[49] On the other hand, authors and creators witness their work being scraped from the open web to train AI models, including GenAI models that produce outputs that compete with their work, undercutting revenue and potentially affecting their livelihood.[50]

---

44 Department for Science, Innovation and Technology, Department for Culture, Media and Sport, & Intellectual Property Office. (2026, March 18). In Copyright and artificial intelligence: Impact assessment (p.12).https://assets.publishing.service.gov.uk/media/69ba68f7c06ba9576435abb0/CP2602959_-_AI_and_Copyright_Impact_Assessment_Web.pdf

45 Widder, D. G., West, S. M., & Whittaker, M. (2023, August 17). Open (for business): Big Tech, concentrated power, and the political economy of open AI (pp. 8–10). SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4543807

46 United States uses the fair use doctrine to allow copyright exemption in certain cases which are determined using the four-factor test: (1) Purpose and character of the use, including whether the use is of a commercial nature or is for nonprofit educational purposes: (2) Nature of the copyrighted work: (3) Amount and substantiality of the portion used in relation to the copyrighted work as a whole; (4) Effect of the use upon the potential market for or value of the copyrighted work. See, https://www.copyright.gov/fair-use/

47 Countries that have adopted fair dealing include the UK, India and Kenya. A key difference from fair use is that fair dealing usually operates on a closed list of permitted exceptions usually for purposes of scientific research, education, fair criticism etc.

48 Karanja, N., & Koros, C. (2022, December 2). The role of copyright law in text and data mining research. InfoJustice. https://infojustice.org/archives/44946

49 Kilic, B. (2026, March 20). The moratorium the AI industry cannot afford to lose. Centre on Knowledge Governance. https://knowledgegov.org/the-moratorium-the-ai-industry-cannot-afford-to-lose/

50 A 2024 global study commissioned by the International Confederation of Societies of Authors and Composers predicts that generative AI will cause a €22 billion cumulative revenue loss for music and audiovisual creators by 2028 due to unlicensed content usage and market substitution. It estimated that widespread use of AI tools and surges in the AI-generated content market could put the revenues of 24% of music creators and 21% of audiovisual creators at risk.,International Confederation of Societies of Authors and Composers. (2024). Global economic study on the impact of artificial intelligence on the music and audiovisual industries (pp. 71, 86). https://www.cisac.org/services/reports-and-research/cisacpmp-strategy-ai-study

Further, AI-generated summaries in search results, even when accurate and with attribution, has been shown to disincentivize users from visiting publishers' websites, reducing website traffic and affecting advertisement or subscription-based revenue models.[51]

Authors and creators witness their work being scraped from the open web to train AI models, including GenAI models that produce outputs that compete with their work, undercutting revenue and potentially affecting their livelihood

Successful resistance to copyright exceptions claimed by developers hinges on multiple factors. For instance, Germany's TDM exception allows for only non-commercial use, while Japan and Singapore permit TDM exception for commercial use as well[52]. However, in the case of AI, determining the commercial nature of TDM becomes complex as demonstrated by the case of LAION-5b, an open source image dataset. In this case, a German court allowed LAION's use of the TDM exception as it provided its datasets free of charge and therefore constituted a step in the process of acquiring scientific knowledge. This was despite the fact that LAION's datasets are also used for training commercial AI models. Determination of non-commercial nature also becomes tricky in cases where an initially non-profit AI entity trains on datasets and later turns into a for-profit corporate (e.g. OpenAI).[53]

While many copyright laws allow copyright holders to opt out of TDM and require webscrapers to respect such opt-outs, in practice the effectiveness of these opt-outs in striking a fair balance has been doubtful. By placing the burden on creators to opt-out, it disadvantages many small creators with limited awareness of opt-out options and weak bargaining power to exercise it meaningfully. Even where creators opt-out, there is an absence of effective means to verify whether their opt-out has been respected in the absence of meaningful transparency obligations requiring AI developers to disclose training data at a granular level.[54] Further, it is seen that technical measures to block web crawlers are not effective in preventing all types of web crawling and in some cases they have been evaded by AI systems.[55]

---

51 Ofcom. (2025, November 4). The era of answer engines: Generative AI's impact on search experiences and online safety (pp. 71, 86). Ofcom. https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/other/the-era-of-answer-engines---discussion-paper.pdf?v=407803

52 A German non-profit which makes open-sourced artificial intelligence models and datasets.LAION. (n.d.). Large-scale artificial intelligence open network. https://laion.ai/

53 Widder, D. G., West, S. M., & Whittaker, M. (2023, August 17). Open (for business): Big Tech, concentrated power, and the political economy of open AI. SSRN.https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4543807; Gioldstein, P. (2024, November 13). https://legalblogs.wolterskluwer.com/copyright-blog/kneschke-vs-laion-landmark-ruling-on-tdm-exceptions-for-ai-training-data-part-1/

54 Department for Promotion of Industry and Internal Trade. (2025, December). Working paper on generative artificial intelligence and copyright (Part 1) (p. 8). DPIIT. https://www.dpiit.gov.in/static/uploads/2025/12/ff266bbeed10c48e3479c941484f3525.pdf

55 Mehrotra, D., & Marchman, T. (2024, June 19). Perplexity is a bullshit machine. WIRED. https://www.wired.com/story/perplexity-is-a-bullshit-machine/

Moreover, opt-out provisions do not address downstream uses: once content is detached from its metadata and transformed, subsequent reuse becomes difficult to trace or control, rendering the loss of control effectively irreversible.[56] And in case the training has already been done, then opt-out cannot be exercised to get the content removed from the dataset as the primary assumption is that all content owners have opted in.[57]

But the more fundamental issue is the fairness of allowing AI training on copyrighted material without consent, attribution, and reciprocity. In *Authors Guild v. Google,*[58] the US Circuit Court held that Google's act of scanning and digitizing millions of books to create a searchable database that showed small snippets of the books constituted fair use. The court reached this conclusion on the ground that Google Books' search and snippet function constituted a transformative work and did not directly substitute the original books. When it comes to the AI learning context, the two US decisions in *Andrea Bartz v. Anthropic*[59] and *Richard Kadrey v. Meta*,[60] delivered by Judge Alsup and Judge Chabbria, respectively, become instructive. Both cases involved scraping shadow libraries for books to create datasets to train their LLM models. In both cases, the court held that use of copyrighted material for training LLMs is highly transformative. Further, in both cases, the market effect was dismissed as there was no evidence to show that the models produced substantially similar work. On this basis, training of AI on a vast corpora of scraped books was held to fall within fair use.

The two cases also raised a crucial question that is highly contested now between copyright holders and AI actors: does the use of copyrighted works to train GenAI, which result in outputs that compete with, even if they are not substantially similar to,[61] human-created works, amount to a form of market harm recognized under copyright law?[62]

---

56 Department for Promotion of Industry and Internal Trade. (2025, December). Working paper on generative artificial intelligence and copyright (Part 1) (p. 8). DPIIT. https://www.dpiit.gov.in/static/uploads/2025/12/ff266bbeed10c48e3479c941484f3525.pdf

57 Rasin, B. (2025, April 18). Opt-out approaches to AI training: A false compromise. Berkeley Technology Law Journal. https://btlj.org/2025/04/opt-out-approaches-to-ai-training/

58 U.S. Court of Appeals for the Second Circuit. (2015, October 16). Authors Guild, Inc. v. Google Inc. https://www.copyright.gov/fair-use/summaries/authorsguild-google-2dcir2015.pdf

59 United States District Court for the Northern District of California. (2025). Bartz et al. v. Anthropic PBC, No. 3:23-cv-04349 (order on summary judgment). https://storage.courtlistener.com/recap/gov.uscourts.cand.434709/gov.uscourts.cand.434709.231.0_3.pdf

60 United States District Court for the Northern District of California. (2025). Kadrey et al. v. Meta Platforms, Inc., No. 3:23-cv-03417 (order on summary judgment). https://storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.598.0_1.pdf

61 A judicial decision in China indicates that it may be easier to establish copyright infringement when the AI generated work that is substantially similar to the copyrighted work that it has been trained on In Shanghai Character License Administrative Co. v. TAB, Guangzhou Internet Court held that a generative AI platform ("Tab") infringed the copyright of Shanghai Character License Administrative Co., Ltd. (SCLA) by enabling users to generate images resembling Ultraman, over which SCLA holds exclusive rights. The court accordingly found the platform liable for copyright infringement. See, https://legalblogs.wolterskluwer.com/copyright-blog/awaken-the-warrior-of-light-what-lessons-can-we-learn-from-ultraman/

62 Agrawal, A. (2025, July 2). [Part II] Obiter > Ratio: The good and bad in the first two American decisions on generative AI and copyright. SpicyIP. https://spicyip.com/2025/07/part-ii-obiter-ratio-the-good-and-bad-in-the-first-two-american-decisions-on-generative-ai-and-copyright.html

While this question was not argued in the Meta case, Judge Chhabria observed that the AI-generated work produced at scale by models trained on copyrighted work, even if not substantially similar, can create indirect competition with human created work, thereby disincentivising creative efforts. On the other hand, this argument was dismissed by Judge Alsup the Anthropic case on the ground that what GenAI produces is a pro-competitive effect, the consequences of which are beyond copyright's remit to address. Hence, displacement of creators and loss of revenue due to competition when there is no substantial similarity between the works is irrelevant to fair use consideration. Judge Alsup explains his reasoning thus:

> ***Claude has outputted grammar, composition, and style that the underlying LLM distilled from thousands of works. But if someone were to read all the modern-day classics because of their exceptional expression, memorize them, and then emulate a blend of their best writing, would that violate the Copyright Act? Of course not. Copyright does not extend to "methods of operation, concept, principles illustrated or embodied in a work.***[63]

Scholars have, however, pointed to some concerns with this human analogy and the risk it seeks to downplay. While both human learning and machine learning involve identification and extraction of knowledge, patterns, styles, and convention, where they differ is the "industrial scale of the progress" of AI training as "no human could read, let alone retain, that volume of material."[64] Further, the amount of new work that a human being produces after learning from existing work is far modest than the millions of outputs that an AI model can produce at negligible cost at "superhuman speed and scale" and which can also pose the real risk of competitive displacement.[65] This raises the question as to whether the established doctrines of copyright law hold well in the AI context.

The issue of competitive displacement has been raised in copyright lawsuits against AI companies in other jurisdictions as well. In Brazil, Folha, a major newspaper company, has filed a lawsuit against OpenAI for copyright infringement and unfair competition. Folha claims unfair competition by Open AI generating summaries of news articles, thereby discouraging users from visiting the newspaper website, leading to loss of traffic, ad revenue and subscriptions.[66]

---

63 United States District Court for the Northern District of California. (2025). Bartz et al. v. Anthropic PBC, No. 3:24-cv-05417 (order on summary judgment) (p. 12).https://storage.courtlistener.com/recap/gov.uscourts.cand.434709/gov.uscourts.cand.434709.231.0_3.pdf

64 Guadamuz, A. (2025). Written evidence submitted to the UK Parliament Committee on artificial intelligence and copyright (AIC0026) (pp. 6–7). https://committees.parliament.uk/writtenevidence/162681/pdf/

65 U.S. Copyright Office. (2025, May 9). Copyright and artificial intelligence: Part 3—Generative AI training (Pre-publication version). https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf

66 Alvarenga, M., Schirru, L., Miceli, E., Gonçalves, L., & Gruber, H. (2025, November 18). The first major generative AI and copyright case in Brazil: First impressions and challenges ahead. Kluwer Copyright Blog. https://legalblogs.wolterskluwer.com/copyright-blog/the-first-major-generative-ai-and-copyright-case-in-brazil-first-impressions-and-challenges-ahead/

Recently, Brazil's competition authority has also launched a probe on Google over the anti-competitive effect of Google's use of news content.[67] In South Korea too, a lawsuit by the country's three terrestrial broadcasters (KBS, MBC, and SBS) against the AI company Naver raises the issue of anti-competitive practices in addition to copyright infringements.[68]

These debates point to the limitations of the copyright regime in safeguarding the economic and moral rights of creators over their works in the AI context. Moreover, copyright law fails to capture the full spectrum of extractive AI practices, as AI systems draw upon a far broader reservoir of information and cultural resources, many of which typically fall outside the ambit of copyright protection. In doing so, they risk undermining the contributions and stewardship of those who produce, curate, and sustain these knowledge commons. This issue is explored further in the next section.

Copyright law fails to capture the full spectrum of extractive AI practices, as AI systems draw upon a far broader reservoir of information and cultural resources, many of which typically fall outside the ambit of copyright protection

4.2. Enclosure through trade secrets and patents

The extraction of copyrighted materials as well as shared knowledge resources and cultural commons in the AI moment is followed by its re-enclosure *via* proprietary models and the abuse of IP to entrench control over the knowledge economy. In this respect, the role of trade secrets and patents comes to the fore.

Concerns have been raised about the increasing use of trade secrets to enclose data undergirding AI systems. The scope of trade secrets is often vague and interpreted differently across jurisdictions. Further, unlike copyright and patents, trade secrets are not time-limited. Together, these factors enabled an overexpansion of trade secret claims to cover a widening range of information goods of the nature of data in both commercial and non-commercial contexts—sometimes even extending to publicly available information. Trade secret claims over datasets have two serious consequences.

---

67 Martins, L. (2026, April 24). Brazil's competition watchdog opens Google probe over publisher pay. Tech Policy Press. https://www.techpolicy.press/brazils-competition-watchdog-opens-google-probe-over-publisher-pay/

68 Lee, H.-r. (2025, January 14). Broadcasters sue Naver for copyright violations. The Korea Times. https://www.koreatimes.co.kr/southkorea/society/20250114/korean-broadcasters-sue-naver-for-copyright-violations; A few months after the lawsuit was filed, KBS subsequently partnered with Naver for AI-driven media solutions. However, the issue is far from resolved as evidenced by the recent lawsuit by South Korea's leading terrestrial broadcasters against OpenAI, claiming that the company trained its ChatGPT model using their news content without permission. See, Digital Watch Observatory. (2026, February 24). OpenAI faces legal action in South Korea from top networks. https://dig.watch/updates/openai-faces-legal-action-in-south-korea-from-top-networks

First, when firms claim trade secrets over social data commons, they enclose what could otherwise be a shared social resource. This leads to a 'tragedy of the anticommons,' thereby stifling innovation as many economic actors are prevented from accessing and leveraging data's full value. Second, trade secret protections reduce transparency and accountability in AI systems built on such data. This can produce harmful outcomes, as individuals cannot scrutinize or challenge algorithmic decisions affecting them (See Box 1)

Box 1

The case of trade secrets claim over privately-mediated data of public importance

The two consequences of overexpansion of trade secret claims are illustrated by the case of *Lyft Inc v City of Seattle*[69], wherein Lyft, a ride hailing service, invoked trade secret claims to avoid submitting quarterly reports—which included various data categories, such as the total number of rides and pick-up and drop-off zip codes—to the city administration. Lyft expressed concerns of maintenance of confidentiality despite the city administration's implementation of measures to safeguard data. This limited the reuse opportunity of such data by public actors.

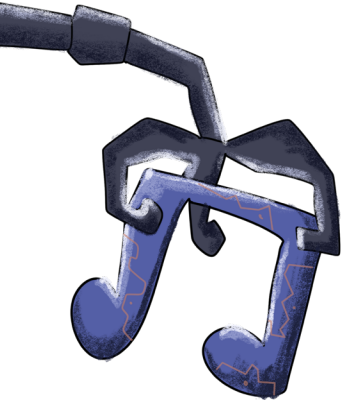
Additionally, when a ride-share analyst filed a request under the Public Records Act to analyse data to see if the ride-sharing service were fairly serving communities of color, the King County Superior Court issued a permanent injunction enjoining the reports from disclosure, maintaining that the zip code reports are trade secrets. But on appeal the Washington Supreme Court reversed this decision and granted access to the report. The court held that the report is a public record even if it is under trade secrets and it can be lawfully withheld only if it was determined that such disclosure "would clearly not be in the public interest and would substantially and irreparably damage a person or a vital government interest."[70] In doing so, the case establishes an important threshold for limiting trade secret claims, preventing their overexpansion and safeguarding the public interest and the rights and liberties of others.

The control of a few corporate actors over future arc of AI innovation is also aided by the using the proprietary measures and the patent regime to monopolize and control access to AI models after they have been built on data constituting knowledge and cultural commons. Evidence from the agricultural sector illustrates this dynamic: as digital agriculture platforms powered by GenAI proliferate, data collected from farmers and farming practices are used to train models that are subsequently enclosed as proprietary systems, often accessible to those very farmers only through paid subscription models.[71]

69 Lyft, Inc v City of Seattle 94026-6 (Wash Sup Ct 2018).

70 Lyft, Inc v City of Seattle 94026-6 (Wash Sup Ct 2018).

71 Rosado, S., & Vogliano, S. (2025, December 5). Commons to code: How platforms rewire agriculture and reshape power. ETC Group https://www.etcgroup.org/content/commons-code-how-platforms-rewire-agriculture-and-reshape-power-0



The control of a few corporate actors over future arc of AI innovation is also aided by the using the proprietary measures and the patent regime to monopolize and control access to AI models after they have been built on data constituting knowledge and cultural commons.

When it comes to genetic and biological resources, the AI context intensifies the longstanding concerns around biopiracy and patents being used to appropriate genetic resources and traditional knowledge without fair recognition or benefit sharing. The emergence of digital sequence information (DSI)—i.e., digitalized versions of genetic sequences—and generative biology further complicates this landscape, as the Convention on Biological Diversity (CBD) and its access and benefit-sharing frameworks are still nascent on the governance of digitized genetic data.[72] DSI is often accessed from public databases like the GenBank database that allow for anonymous access to genetic data with no strings attached.[73] AI enabled generative biology products, built on DSI, have raised concerns of living materials sourced from communities being transformed into data and converted into proprietary assets through dense patent thickets. This undermines equitable benefit sharing obligations, restricts public access and reinforces asymmetries in control over both innovation and biological knowledge.[74] In essence, as a participant in the Ibero-American FGD observed, the cultural and biological origins of data are obscured once they are absorbed into the AI circuits; communities lose control over their biocultural heritage and the “code of life” is privatized.[75]

---

72 Generative Biology (GenBio) describes the application of generative artificial intelligence (AI that creates novel or synthetic digital material) to the field of genetic engineering. See, Thomas, J. (2025, December 5). AI’s large looting models? The emerging generative biology stack as the next frontier of biopiracy. ETC Group. https://www.etcgroup.org/content/commons-code-how-platforms-rewire-agriculture-and-reshape-power-0

73 Rao, C. & Gopakumar, K.M. (2024). Exclusion of DSI undermines the effectiveness of WIPO’s Proposed International Legal Instrument Relating to Intellectual Property, Genetic Resources and Traditional Knowledge. Third World Network. https://www.twn.my/announcement/WIPO%20IGC%20-%20DSI%20%20working%20document_21052024.pdf

74 Thomas, J. (2026, forthcoming). Large Looting Models: When AI Theft meets Biopiracy. In Knowledge Futures. Bot Populi. https://botpopuli.net/section/knowledge-futures/; Thomas, J. (2026, February 17). IP Regimes, mainstream AI Innovation models, and implications for the public domain. In Cultural Rights, Innovation, and Development in the AI moment: Towards a public domain framing. IT for Change. https://itforchange.net/event-report-cultural-rights-innovation-and-development-ai-moment-towards-a-public-domain-framing/

75 Vogliano, S. (2026, focus group discussion). Oral inputs at the regional focus group discussion- Ibero-America: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 16 2026)

Box 2

The case of Stevia

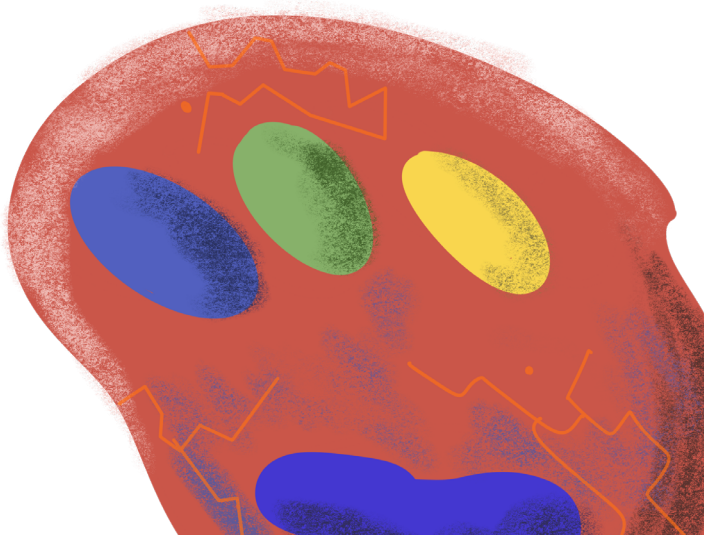
Stevia is a plant-based sweetener used in foods and beverages, extracted from the leaves of Stevia rebaudiana, a plant native to the Paraguay–Brazil border region in South America. It is often promoted as a zero-calorie substitute for sugar.

A research study by ETC Group[76] notes that companies like Arzeda have used AI to engineer enzymes that convert common stevia compounds into more commercially desirable variants, now marketed by brands such as Splenda as "natural." These lab-derived products directly compete with farmer-grown stevia cultivated by Indigenous communities like the Guaraní people in Paraguay, and threaten to drive down prices for small producers. Beyond economic impacts, this shift undermines biocultural rights; as ETC Group notes, "when AI-driven GenBio systems enclose these resources and transform them into proprietary digital assets, they undermine sovereignty over genetic resources and sever cultural relationships to biodiversity."[77]

To sum up, on one hand, IP frameworks like copyright permit the large-scale extraction of creative works while remaining structurally ill equipped to address the market-substitution harms that creators increasingly face. At the same time, IP such as trade secrets and patents enable firms to enclose and exert proprietary control over AI systems built on this extracted corpus of creative work as well as the knowledge and cultural commons. The black-box nature of AI becomes a red herring to ward off compensation claims and accountability measures. Thus, what was once enclosed is also now untraceable.

76 Thomas, J. (2025, December 5). AI's large looting models? The emerging generative biology stack as the next frontier of biopiracy. ETC Group. https://www.etcgroup.org/content/commons-code-how-platforms-rewire-agriculture-and-reshape-power-0

77 Thomas, J. (2025, December 5). AI's large looting models? The emerging generative biology stack as the next frontier of biopiracy. ETC Group. https://www.etcgroup.org/content/commons-code-how-platforms-rewire-agriculture-and-reshape-power-0



5. Contradictions in Intellectual Property Frameworks for AI

Section summary

AI introduces a new paradigm of creative and knowledge production that challenges the conceptual limits of IP frameworks in safeguarding economic and moral rights of creators.

- The opacity and scale of AI systems make it difficult to trace individual creative works once they are tunneled into AI training systems.
- AI does not compete work-for-work, but generates competitive creative outputs at lower costs by abstracting patterns and styles across large datasets of creative works. This threatens displacement of creative labor and value erosion, a challenge that copyright frameworks are ill equipped to address.
- If the lowering of the costs of creative innovation due to AI undermines the traditional economic rationale for copyright protection, a similar logic applies in the patent regime. The lowering of production and innovation costs through AI offers a reason to reconsider the scope and strength of patent claims over AI-driven inventions.
- More broadly, the individualistic and propertarian moorings of IP frameworks renders them ill equipped to protect the collective, intergenerational, and culturally embedded forms of knowledge that are increasingly under pressures of dispossession and commodification.

The extractivist tendency of the dominant AI paradigm raises questions beyond IP protection and compensation of individual creators. Nevertheless, the debate often centers around expanding and adapting conventional IP protections like copyright to the AI context, treating them as a panacea for addressing both the market as well as non-market consequences of large-scale extraction and enclosure of creative works.

This section examines the inherent limitations and contradictions of conventional IP frameworks, especially copyright, when applied to AI context and how the propertarian logic of IP renders itself inadequate, and at best, a partial response to the structural shifts in creative and cultural production due to AI.

5.1. A category error: Limits of copyright in the AI context

AI presents challenges to existing IP frameworks, because it introduces a new production paradigm of knowledge products that does not appear to rely on traditional incentives for creativity. For instance, copyright—the main IP in question in policy discussions on AI extractivism—was developed to govern individual creative works and protect against their unauthorized copying and reproduction. This is because as creative endeavour is expensive while copying is cheap, authors need to be protected against being undercut by cheap copies of their own work. The underlying logic is one of economic efficiency by balancing creator incentives against public access to maximize public good.

If this is the rationale of copyright law, then it falls apart in the AI context. AI systems do not simply copy or reproduce individual works. Rather, they rely on the large-scale aggregation of data from diverse sources—much of it copyrighted—from which statistical patterns, relationships, and stylistic features are extracted after converting them into machine readable or tokenized formats. As a result, individual pieces of work are no longer recognizable in their identifiable expression and any single piece carries negligible value in relation to the whole dataset.[78] This, combined with the opacity surrounding training datasets, makes it exceedingly difficult to trace outputs back to specific works or authors. Consequently, ensuring attribution and mounting a viable copyright infringement claim, becomes structurally difficult. Even licensing arrangements and determination of fair compensation becomes difficult at the training stage due to uncertainty about the commercial potential of the model at the stage of deployment and given the possibility of long-term reuse.

At the output stage of AI, copyright infringement again becomes difficult to establish as in the majority of cases, “these outputs may be considered novel compositions that synthesize patterns derived from many different sources.”[79] Thus, the issue is not just a simple pro-competitive effect as Judge Aslup puts it in the Meta case; it is fundamentally a threat to displace the human role in cultural production itself and a labor market transformation.

---

78 House of Lords Communications and Digital Committee. (2026, March 6). AI, copyright and the creative industries (4th Report of Session 2024–26). https://publications.parliament.uk/pa/ld5901/ldselect/ldcomm/267/267.pdf

79 Even in cases where the output may have substantial similarity with the copyrighted work where copyright law enforcement should be relatively easier, in practise, it is often prohibitively difficult for rightsholders to detect, evidence and enforce such claims. House of Lords Communications and Digital Committee. (2026, March 6). AI, copyright and the creative industries (4th Report of Session 2024–26, HL Paper 267, pp. 26–27). https://publications.parliament.uk/pa/ld5901/ldselect/ldcomm/267/267.pdf

Even in cases where copyright law provides exceptions to allow AI training in the name of fair use or TDM mining, it does not have mechanisms to "ensure economic sustainability of information production within those permitted uses."[80] These are structural issues that the established doctrine and principles of copyright protection are ill equipped to address. Instead, as discussed earlier, copyright often becomes a tool for consolidating control in the hands of a few dominant players, reinforcing market concentration, and limiting broader participation in digital and creative economies as discussed in Section 3.2. Thus, the anachronism of conventional IP frameworks in relation to the AI context, coupled with an overreliance on them to address the questions of AI extractivism, is proving to be increasingly costly for rightsholders. As Akshat Agrawal argued, "we may be committing a category error by using the copyright framework to respond to challenges raised by the AI moment."[81]

Thus, the issue is not just a simple pro-competitive effect as Judge Aslup puts it in the Meta case; it is fundamentally a threat to displace the human role in cultural production itself and a labor market transformation.

If the decreasing cost of creation makes copyright infringement claims against AI difficult to sustain, then extending this logic to the patent regime raises similar questions. As discussed during the roundtable, to the extent that patent monopolies are traditionally justified as incentives for costly innovation, the growing ability of AI systems to reduce the costs of innovation weakens this rationale. In fields such as synthetic biology, this creates a compelling opening to reconsider and recalibrate the scope and strength of patent protection for AI-driven innovations.[82]

5.2. Misalignment with collective knowledge systems

The limitations of current IP frameworks in addressing the issues arising from an extractive AI paradigm also stem from the individualistic and propertarian moorings of IP, which make them ill equipped to protect the collective, intergenerational, and culturally embedded forms of creativity that underpin many forms of Indigenous and traditional knowledge systems and shared resources.

80 Keller, P. (2025, June 30). Beyond AI and copyright: Funding a sustainable information ecosystem (p. 8). Open Future. https://openfuture.eu/wp-content/uploads/2025/06/250630_Beyond-AI-and-copyright-funding-a-sustainable-information-ecosystem.pdf

81 Agrawal, A. (2026, forthcoming). Beyond Copyright: Why IP regimes cannot address AI's displacement of creative labor. In Knowledge Futures. Bot Populi https://botpopuli.net/section/knowledge-futures/; Agrawal, A. (2026, February 17). IP Regimes, mainstream AI Innovation models, and implications for the public domain. In Cultural Rights, Innovation, and Development in the AI moment: Towards a public domain framing. IT for Change. https://itforchange.net/event-report-cultural-rights-innovation-and-development-ai-moment-towards-a-public-domain-framing/

82 Barooah, S. P. (2026, forthcoming). AI and IP Knowledge Governance Clashes. In Knowledge Futures. Bot Populi https://botpopuli.net/section/knowledge-futures/; Barooah, S. P., & Gopakumar, K. M. (2026, February 17). Legal-policy pathways to enrich the data and AI commons. In Cultural Rights, Innovation, and Development in the AI moment: Towards a public domain framing. IT for Change. https://itforchange.net/event-report-cultural-rights-innovation-and-development-ai-moment-towards-a-public-domain-framing/

Indigenous communities and their cultural rights are uniquely vulnerable to harm in the emerging AI landscape because these technologies frequently extract and commodify their cultures, values, knowledge, narratives, aesthetics, and diverse artistic expressions.[83] As participants in the Asia-Pacific roundtable observed, even those IP frameworks that guarantee collective rights like Geographical Indications have often had limited practical effect in protecting communities, and are poorly suited to enforce claims against AI-driven appropriation.[84]

It is not only collective cultural expressions that are at risk of intensified commodification in the AI age, but also genetic and biological resources. A recent study in the context of the rise of AI-driven Generative Biology (GenBio) models suggests that in most cases, relevant rightsholders such as Indigenous People, local communities, and farmers are not aware of how biological and genetic samples are collected from their territories or bodies and fed into AI models fuelling the GenBio industry.[85] This violates the right of Indigenous Peoples to maintain, control, protect, and develop their cultural heritage, human and genetic resources, seeds, medicines, knowledge of the properties of fauna and flora, etc., and the manifestations of their science, technologies, and cultures—a right protected under Article 31 of the UN Declaration on Indigenous People's Rights.

The current IP-AI debate tends to narrow its focus on compensation for rightsholders. It sidesteps the fact that the vast training corpus of AI also includes resources in public archives, libraries, and cultural heritage institutions, linguistic data, open content projects like Wikimedia, open science initiatives, government and public data, and even any form of digital cultural data generated through intentional or unintentional acts of cultural expression—most of which operate outside this logic of compensation.[86] However, all of these are now increasingly being folded into proprietary systems that consolidate the control of a few corporate actors over cultural production and dissemination, without commensurate benefits flowing back to the commons. This dynamic can be illustrated by taking the case of language datasets, often drawn from national libraries and archives. The development of foundational models capable of supporting linguistic diversity, especially low-resource languages, has become a competitive flex for large AI companies.[87]

---

83 United Nations General Assembly. (2025, July 30). Artificial intelligence and creativity: Report of the Special Rapporteur in the field of cultural rights (A/80/278). https://www.ohchr.org/en/documents/thematic-reports/a80278-artificial-intelligence-and-creativity-report-special-rapporteur ; Hellmann, O. (2025, October 24). Historical images made with AI recycle colonial stereotypes and bias – new research. The Conversation. https://theconversation.com/historical-images-made-with-ai-recycle-colonial-stereotypes-and-bias-new-research-268070

84 Sangma, N. & Christine, D.I. (2026, March 10). Oral inputs at the regional focus group discussion- Asia-Pacific: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 10 2026).

85 Thomas, J. (2025, December 5). AI's large looting models? The emerging generative biology stack as the next frontier of biopiracy. ETC Group. https://www.etcgroup.org/content/commons-code-how-platforms-rewire-agriculture-and-reshape-power-0

86 UNESCO. (2025, September 23). Report of the Independent Expert Group on artificial intelligence and culture. https://www.unesco.org/sites/default/files/medias/fichiers/2025/09/CULTAI_Report%20of%20the%20Independent%20Expert%20Group%20on%20Artificial%20Intelligence%20and%20Culture%20%28final%20online%20version%29%201.pdf

87 Vincent, J. (2022, November 2). Google plans giant AI language model supporting world's 1,000 most spoken languages. The Verge. https://www.theverge.com/2022/11/2/23434360/google-1000-languages-initiative-ai-llm-research-project; Vincent, J. (2022 July 6). Meta open sources early-stage AI translation tool that works across 200 languages. The Verge. https://www.theverge.com/2022/7/6/23194241/meta-facebook-ai-universal-translation-project-no-language-left-behind-open-source-model ; Shivam, H. (2025, November 12) Meta's new Omnilingual AI can understand 1,600+ languages: How it works. Business Standard. https://www.business-standard.com/technology/tech-news/meta-s-new-omnilingual-ai-can-understand-1-600-languages-how-it-works-125111200466_1.html

Governments are also partnering with the private sector to digitize records, libraries, and archives and use them to create language datasets.[88] While these open up possibilities for the preservation of linguistic and cultural resources and for more inclusive AI systems, researchers have pointed out that the resulting language datasets are rarely hosted openly or made accessible to the broader public, despite being built from commons-based knowledge resources. What this brings to fore is not just concentration of control over innovation possibilities in a few hands, but also difficulties in preserving and maintaining commons resources for collective benefit. Since IP frameworks gloss over the incremental and collective nature of knowledge production, they remain ill-equipped to ensure the long-term sustainability of information and cultural creation. Such sustainability requires not only attribution-based compensation, but also mechanisms to redistribute value generated back to the commons. Further, as a participant in the Africa FGD noted, reciprocity of benefits cannot be a one-time act; it must be continuous and commensurate with value generated over time through the further development and commercialization of AI models that rely on datasets derived from the commons.[89]

Since IP frameworks gloss over the incremental and collective nature of knowledge production, they remain ill-equipped to ensure the long-term sustainability of information and cultural creation.

Further, the shift toward AI-mediated access to information risks making traditional knowledge institutions such as libraries, archives, publishers, and cultural organizations less visible and potentially obsolete, undermining both the public relevance and funding of these institutions.[90] In this scenario, as Paul Keller argues, “the challenge is not simply to protect the rights of individual creators and their creative industry intermediaries, but to ensure the continued vitality of the public infrastructures that support knowledge production, preservation, and access and the public goods that result from them.”[91] This raises questions of control over data, algorithms, and AI infrastructures and how it can be reasserted for the public good. These questions need policy mechanisms beyond IP.

---

88 Chandrasekhar, R. (2025, March 10). Legal frictions for data openness: The open web and AI. ODECO. (pp 19). https://ok.hypotheses.org/files/2025/03/Legal-frictions-for-data-openness-open-web-and-AI-RC-2025-final.pdf

89 Ndubane, D. (2026, focus group discussion). Oral inputs at the regional focus group discussion-Middle East & Africa: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 11 2026).

90 Keller, P. (2025, June 30). Beyond AI and copyright: Funding a sustainable information ecosystem (pp. 4–5). Open Future. https://openfuture.eu/wp-content/uploads/2025/06/250630_Beyond-AI-and-copyright-funding-a-sustainable-information-ecosystem.pdf

91 Keller, P. (2025, June 30). Beyond AI and copyright: Funding a sustainable information ecosystem (p. 9). Open Future. https://openfuture.eu/wp-content/uploads/2025/06/250630_Beyond-AI-and-copyright-funding-a-sustainable-information-ecosystem.pdf

6. Reclaiming the knowledge infrastructures of tomorrow

Section summary

Addressing AI extractivism requires moving beyond protecting outputs or compensating individual creators towards governance models that centre cultural rights, community control, and the sustainability of collective knowledge systems.

- Cultural rights must ensure that knowledge-producing communities retain meaningful control over how their knowledge is represented, transformed and used within AI systems.
- A rights-based understanding of “living culture," foregrounding knowledge as a collective and evolving process, is central to this shift.
- Despite the structural limitations of IP frameworks in addressing AI’s impact on the cultural and knowledge commons, certain IP reforms are necessary to bring a stronger focus on equity and creativity as opposed to just economic incentives.
- Other policy pathways include strengthening community data governance through stewardship-based approaches; building public data and AI infrastructures to counter the concentration of knowledge infrastructure and epistemic power; and mechanisms to redistribute value, including fiscal levies, to help sustain the knowledge ecosystems on which AI development depends.

The preceding analysis underscores the need for a fundamental shift in how AI infrastructures are developed and governed. The current paradigm—marked by large-scale extraction of cultural and informational resources and their incorporation into proprietary systems—places sustained pressure on the cultural commons and constrains the conditions under which the public domain can be replenished. At stake is not only the appropriation of discrete works, but a broader reorganization of cultural production, in which shared knowledge systems and everyday cultural practices are continuously captured and repurposed as inputs into AI systems.

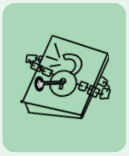
Reclaiming the knowledge infrastructures of tomorrow therefore requires moving beyond extractive models towards approaches that are collaborative and accountable to the communities whose knowledge systems they draw upon.

Potential governance pathways to reclaim the knowledge infrastructures of tomorrow

- Shift to a rights-based understanding of “living culture” — i.e., culture as an ongoing, collective process of knowledge production and meaning-making rather than a static repository of works.
- Ensure meaningful control of knowledge-producing communities over how their knowledge is represented, transformed, and used within AI systems.

Intellectual Property reforms: *Necessary but insufficient*

Thinning trade secrets claim



- Balance trade secret protections with transparency and accountability obligations in the public interest.
- Prevent the 'tragedy of the anticommons' that stifles innovation and access to knowledge.

Promoting social and collective licensing frameworks

- Explore alternative licensing approaches, including community-centred and reciprocity-based arrangements.
- Develop tiered licensing approaches responsive to different forms of cultural and community knowledge (for example, Nwulite Obodo Open Data License).

Beyond IP: *Governing knowledge infrastructures*

Strengthening community data governance

- Operationalize stewardship-based approaches to data governance.
- Develop public data infrastructures that enable accountability and collective oversight.
- Explore the development of a global AI training data framework with baseline obligations on transparency, provenance, and benefit-sharing.

Building public AI infrastructure

- Invest in compute resources, datasets, models, and institutional capacity.
- Support the development of public AI ecosystems with sustained public investment and democratic governance arrangements.

Redistributing value

- Develop mechanisms to redistribute value generated from publicly available cultural and informational resources.
- Consider fiscal levies on commercial AI services trained on publicly available data to support public-interest institutions, cultural ecosystems, and commons-based infrastructures.

A central element of this shift is a rights-based understanding of “living culture”, which recognizes culture as an ongoing, collective process of knowledge production and meaning-making rather than a static repository of works.[92] Preserving cultural rights in the context of AI must therefore extend beyond protecting outputs or compensating individual creators, to ensuring that knowledge-producing communities retain meaningful control over how their knowledge is represented, transformed and put to use within AI systems.

A central element of this shift is a rights-based understanding of “living culture”, which recognizes culture as an ongoing, collective process of knowledge production and meaning-making rather than a static repository of works.

How such control is distributed also shapes the vitality of shared knowledge ecosystems, including the cultural commons. Where communities are able to influence the terms on which their knowledge enters and circulates within these ecosystems, participation can be sustained in ways that support continued contribution and regeneration. In contrast, where incorporation takes place without such influence, the commons risks becoming a site of one-way extraction, weakening the incentives and conditions for knowledge production and sharing. Moreover, AI systems ingest and repackage publicly available information, including collections curated by public and cultural institutions, while simultaneously displace these institutions as primary sites of access.[93] This risks eroding both the engagement of users with and funding incentives for these institutions, thereby undermining the infrastructures that sustain the cultural commons and public domain.

These concerns are closely tied to broader questions of development. As Amartya Sen has argued, development is fundamentally about expanding capabilities, including the ability to participate in and shape processes of knowledge production.[94] The governance of AI infrastructures therefore has direct implications on who can build and exercise these capabilities.

Reforming IP frameworks is a necessary intervention, since they currently enable the large-scale extraction and reuse of cultural resources within AI systems, often without meaningful consent or compensation, as outlined in Section 3.

---

92 IT for Change, & Center for Global Digital Justice. (2025, December). Joint submission to the call for input for the EMRTD study “Artificial intelligence, cultural rights, and the right to development”. IT for Change. https://itforchange.net/joint-submission-to-call-for-input-for-emrtd-study-%e2%80%98artificial-intelligence-cultural-rights-and/

93 Keller, P. (2025, June 30). Beyond AI and copyright: Funding a sustainable information ecosystem. Open Future. https://openfuture.eu/wp-content/uploads/2025/06/250630_Beyond-AI-and-copyright-funding-a-sustainable-information-ecosystem.pdf

94 Sen, A. (1999). Development as Freedom. Oxford University Press.

The UNESCO Recommendation on the Ethics of AI also recommends fostering research at the intersection between AI and IP to assess how AI technologies are affecting the rights or interests of IP owners, whose works are used to research, develop, train or implement AI applications.[95] At the same time, as the following section shows, IP reform can only partially address these dynamics, given the broader political-economic conditions under which AI infrastructures currently operate.

6.1. IP reform: Necessary but insufficient

Despite the IP regime's conceptual and structural limitations in addressing the impact of AI extractivism's on the cultural and knowledge commons, some reform pathways are important to explore. Such reforms are necessary to check against IP frameworks becoming vehicles of extraction and commodification, and to bring a stronger focus on equity and creativity as opposed to narrow economic incentives.

Thinning trade secret claims: IP protections like trade secrets should not become a hindrance to data accessibility and transparency of AI systems. National courts and parliaments should stymie the over-expansion of trade secret claims in non-personal or aggregate data than what is permissible under TRIPS agreement and national frameworks. This is crucial to balance the rights of non-owners of data and to safeguard freedom of information, freedom of scientific research, and freedom of free movement of data to enable the provision of welfare services and allow for fair remuneration mechanisms.[96] Trade secrets should also be appropriately balanced with transparency and accountability to safeguard public interest.[97] Instructive in this regard is the European Court of Justice' statement that trade secrets or IP cannot be used as a ground by companies to argue for non-disclosure of their algorithms to explain AI systems.[98] Such an approach in trade secret claims adjudication is important to "eliminate the reliance on trade secrets to restrict data access, except for very few (judicially identified) considerations."[99]

Community/social licensing framework: Equitable or community-driven licensing frameworks have been proposed as a potential way to redistribute value back to the communities that sustain the common.[100] Open/permissive licensing frameworks, which were devised by volunteers, non-profits and socially conscious individuals and communities to mitigate the restrictiveness of the copyright regime and enable the reuse of protected work, have faltered in the AI context.

95 UNESCO. (2021, November 23). Recommendation on the ethics of artificial intelligence (para. 99). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137

96 Fia, T. (2022, May 1). Resisting IP overexpansion: The case of trade secret protection of non-personal data. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4123311

97 BRICS. (2025, July 7). "Artificial intelligence must not be a privilege for the few, nor a tool of manipulation in the hands of millionaires," declared Lula at the BRICS Summit. BRICS. https://brics.br/en/news/artificial-intelligence-must-not-be-a-privilege-for-the-few-nor-a-tool-of-manipulation-in-the-hands-of-millionaires-declared-lula-at-the-brics-summit

98 Stankovich, M. (2023). Global toolkit on AI and the rule of law for the judiciary. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000387331/PDF/387331eng.pdf.multi

99 Gurumurthy, A., Sen, S., & Rielli, M. (2024, June). Private algorithms and public interest: Overhauling the trade secrets regime for equitable AI futures (p. 5). T20 Brazil. https://www.t20brasil.org/media/documentos/arquivos/TF05_ST_05_Private_Algorithms66cf69c5d2971.pdf

100 Chandrasekhar, R. (2025, March 10). Legal frictions for data openness: The open web and AI. ODECO. (pp. 44–50). https://ok.hypotheses.org/files/2025/03/Legal-frictions-for-data-openness-open-web-and-AI-RC-2025-final.pdf

These licenses are frequently exploited by well-resourced actors like OpenAI to extract value from the shared informational resources without contributing value back to the commons as required by the licensing conditions. Enforcement also becomes tricky due to the difficulty of tracking the traces of licensed artefacts across the AI lifecycle.[101]

On the other extreme, some jurisdictions have proposed mandatory blanket licence for AI training using lawfully accessed copyrighted works, coupled with a statutory remuneration right for copyright holders.[102] Such systems are designed to ensure wider access to training data, reduce transaction costs, prevent bias and hallucinations in AI systems, and guarantee fair compensation through a centralized, government-designated collective mechanism. However, these proposals have drawn severe criticism for undermining creators' autonomy due to the absence of opt-in or opt-out provisions, offering little scope for meaningful compensation due to a flat, revenue-linked remuneration model, and for going against "voluntary sharing" and "share-alike" principles that underpin many open licensing frameworks.[103]

It is against this context that alternative licensing frameworks that are community-centered are emerging and gaining traction. Instead of one-size-for-all blanket licensing, these frameworks adopt a tiered-approach to preserve greater agency for the communities to decide who can use their licensed IP artefact, for what purposes, and under what conditions for each licensee. An illuminating example in this regard is the *Nwulite Obodo* Open Data License for sharing African datasets, which was also cited in the Africa FGD.[104] Under this model, licensees from within Africa and developing countries receive a royalty-free copyleft license, whereas licensees from developed countries are required to either pay royalties or provide non-monetary benefits, such as commitments to share-alike.

Instead of a one-size-for-all blanket licensing, these frameworks adopt a tiered-approach to preserve greater agency for the communities to decide who can use their licensed IP artefact, for what purposes, and under what conditions for each licensee.

101 Chandrasekhar, R. (2025, March 10). Legal frictions for data openness: The open web and AI. ODECO. (pp. 44–50). https://ok.hypotheses.org/files/2025/03/Legal-frictions-for-data-openness-open-web-and-AI-RC-2025-final.pdf

102 Department for Promotion of Industry and Internal Trade. (2025, December). Working paper on generative artificial intelligence and copyright (Part 1): One nation, one license, one payment—Balancing AI innovation and copyright. DPIIT. https://www.dpiit.gov.in/static/uploads/2025/12/ff266bbeed10c48e3479c941484f3525.pdf

103 Software Freedom Law Center, India. (2026, February 9). Comments on DPIIT working paper on generative AI and copyright (Part 1). SFLC. https://sflc.in/comments-on-dpiit-working-paper-on-generative-ai-and-copyright-part-1/

104 Nzuki, C. (2026, March 11). Oral inputs at the regional focus group discussion-Middle East & Africa: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 11 2026); See Data Science Law Lab's Nwulite Obodo open data license. Data Science Law Lab https://datasciencelawlab.africa/nwulite-obodo-open-data-license/

Another note-worthy example is Kaitiakitanga License, an Indigenous data license created by *Te Hiku* collective in New Zealand based on the Māori data sovereignty principles. The license departs from ownership-based models by recognizing communities or representative non-profits as stewards of their data. The license can prohibit commercial use, but can also be modified to permit it on the condition that profits derived from community data are shared back with the community as royalties.[105] In cases where for-profit LLM developers use the licensed datasets, Te *Hiku* Media recommends that profits generated from such use be shared with the communities instead of being paid to the corporate shareholders.[106]

These alternative licensing frameworks thus seek to overcome the limits of traditional licenses to ensure equitable benefit sharing and greater autonomy of communities over their datasets. However, key challenges remain. These include concerns about its effective enforcement in the absence of supportive legal frameworks; the difficulty of tracing licensed artefacts across the AI lifecycle; and the potential fragmentation of shared understandings of the terms of reuse and the baseline of permissible conduct that conventional open licenses sought to achieve.[107] Addressing these concerns is essential to ensure that the proliferation of licensing frameworks does not once again advantage large commercial actors or enable new forms of data extractivism.

6.2. Beyond IP: Governing knowledge infrastructures

The informational base of AI systems also extends well beyond copyrighted works, drawing on the digital and cultural commons, the public domain, publicly funded research, government data, and decades of everyday cultural expression.[108] Value is generated across these distributed sources but captured within a relatively small number of technological systems. Attempts to address this through work-level attribution or licensing struggle to reflect this structure, given the scale and opacity of AI systems. This points to the need for responses that operate beyond IP, focusing on how knowledge systems are organized, governed, and sustained. As Vipra notes, the task is to recognize “the collective nature of intellectual production” and develop institutions capable of sustaining it accordingly.[109]

---

105 Te Hiku Media. (n.d.). Data sovereignty and the kaitiakitanga license. Tehiku. https://tehiku.nz/te-hiku-tech/te-hiku-dev-korero/25141/data-sovereignty-and-the-kaitiakitanga-license

106 Mahelona, K. (2024, November 8). A practical guide to creating your own stewardship license. Papa Reo. https://blog.papareo.nz/a-practical-guide-to-creating-your-own-stewardship-license/

107 Chandrasekhar, R. (2025, March 10). Legal frictions for data openness: The open web and AI. ODECO. (44–50).https://ok.hypotheses.org/files/2025/03/Legal-frictions-for-data-openness-open-web-and-AI-RC-2025-final.pdf

108 Huang, S., & Siddarth, D. (2023, March 20). Generative AI and the digital commons. arXiv. https://arxiv.org/abs/2303.11074 ; Keller, P. (2026, forthcoming). Who Pays for the Commons? Three Conditions for an AI Levy That Works. In Knowledge Futures. Bot Populi https://botpopuli.net/section/knowledge-futures/; Chandrasekhar, R. (2026, forthcoming). Nurturing AI commons: Reflections from emerging licensing initiatives for training datasets. In Knowledge Futures. Bot Populi https://botpopuli.net/section/knowledge-futures/; Chandrasekhar, R., & Keller, P. (2026, February 17). Legal-policy pathways to enrich the data and AI commons. In Cultural Rights, Innovation, and Development in the AI moment: Towards a public domain framing. IT for Change.

109 Vipra, J. (2026, forthcoming). Against an intellectual property battle over AI: social governance for social production. In Knowledge Futures. Bot Populi. https://botpopuli.net/section/knowledge-futures/

6.2.1. Strengthening community data governance

A central limitation of current approaches lies in how data is treated—as a resource that can be aggregated and reused once accessed, with limited regard for the conditions under which it is produced. This enables large-scale incorporation of cultural and informational resources into AI systems without mechanisms for participation or benefit-sharing, especially where such data reflects collectively produced knowledge.

As participants in the Africa and Asia-Pacific FGDs also emphasized, addressing these dynamics requires greater attention to stewardship-based approaches to data governance. These are premised on the recognition that data is produced through social relations and shared knowledge practices, and therefore, requires governance that ensures accountability towards the communities from whom data is derived.[110] The CARE Principles for Indigenous Data Governance (Collective Benefit, Authority to Control, Responsibility, and Ethics), developed by a coalition of Indigenous data sovereignty networks to complement existing stewardship approaches such as the FAIR Guiding Principles,[111] emphasize collective authority and self-determination of Indigenous peoples in relation to community-related data.[112] Such approaches are also aligned with the 2003 Convention for the Safeguarding of the Intangible Cultural Heritage, which emphasizes the central role of communities in the identification, safeguarding, management, and transmission of cultural heritage.[113]

Stewardship-based approaches to data governance... are premised on the recognition that data is produced through social relations and shared knowledge practices, and therefore, requires governance that ensures accountability towards the communities from whom data is derived.

110 Viljoen, S. (2021). Democratic data: A relational theory for data governance. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3727562

111 Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. Scientific Data, 3, Article 160018. https://doi.org/10.1038/sdata.2016.18

112 Global Indigenous Data Alliance. (n.d.). CARE principles for Indigenous data governance. GIDA. https://www.gida-global.org/care; Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE principles for Indigenous data governance. Data Science Journal, 19(1), Article 43, 1–12. https://doi.org/10.5334/dsj-2020-043 ; Vidotti, S. A. B. G., Torino, E., & Coneglian, C. S. (2024). BeFAIRandCARE: FAIR and CARE principles in research data management. Universidade Tecnológica Federal do Paraná. https://riut.utfpr.edu.br/jspui/bitstream/1/35737/1/faircareprinciplesresearchdata.pdf

113 UNESCO. (2003). Convention for the safeguarding of the intangible cultural heritage (arts. 11 & 15). UNESCO. https://ich.unesco.org/en/convention

The Abalobi initiative in South Africa offers an example of how stewardship-based approaches can be operationalized. It enables small-scale fisher communities to collect and manage data on catch, pricing, and distribution through a co-designed and collectively governed digital platform.[114] This allows communities to retain control over how their data is used, strengthens their bargaining power in markets, and links data governance directly to local resource management. Similarly, *Mukurtu* CMS is an open source system built with Indigenous communities to support them in managing digital cultural heritage through protocols that govern access and circulation.[115] It enables communities to embed culturally-specific permissions, such as restrictions based on kinship, gender or context, directly into the technical infrastructure, ensuring that control over community knowledge persists after its integration into digital systems.

Emerging licensing-based approaches, including reciprocity-based or anti-extractivist licenses, also reflect attempts to strengthen community control. However, as discussed in previous sections, their effectiveness in the context of AI systems remains uncertain. Licensing frameworks operate at the level of individual works or datasets, and are difficult to reconcile with the scale and aggregation involved in AI training, where value is generated across large and heterogeneous data inputs.

This brings into focus the role of institutional design in shaping data governance. In particular, the development of public data infrastructures is a crucial pathway to support commons-based governance in practice.[116] Such infrastructures can enable interoperability, enforce conditions of use, and support forms of collective oversight that are difficult to achieve through contractual or licensing approaches alone. Public data infrastructures can provide the technical and institutional backbone for commons-based governance, enabling data to circulate under terms aligned with public interest objectives while maintaining accountability to the communities from which it originates.

At the same time, the cross-border nature of AI development, where training datasets are assembled and deployed across jurisdictions, raises questions about the adequacy of national-level interventions.[117] In this context, proposals for a global AI training data framework have gained increasing relevance towards ensuring that data is governed in the public interest.[118]

---

114 ABALOBI. (n.d.). ABALOBI: Elevating small-scale fisheries through data and technology. ABALOBI. https://abalobi.org/

115 Washington State University Libraries. (n.d.). Mukurtu CMS. https://mukurtu-australia.libraries.wsu.edu/mukurtu-cms

116 Keller, P. (2026, April 2). Sovereign data commons and public data infrastructure. Open Future. https://openfuture.eu/wp-content/uploads/2026/04/260402_Sovereign-Data-Commons-and-Public-Data-Infrastructure.pdf

117 Participant. (2026, focus group discussion). Oral inputs at the regional focus group discussion- Middle East & Africa: Inputs on cultural rights, innovation, and development in the AI moment: Towards a public domain framing (March 30 2026)

118 IT for Change, & Center for Global Digital Justice. (2025, December). Joint submission to the call for input for the EMRTD study "Artificial intelligence, cultural rights, and the right to development". IT for Change. https://itforchange.net/joint-submission-to-call-for-input-for-emrtd-study-%e2%80%98artificial-intelligence-cultural-rights-and/

Such a framework could establish baseline obligations around dataset transparency,[119] common standards regarding AI data provenance and use,[120] mechanisms for equitable benefit-sharing, and forms of preferential access to communities whose data has contributed significantly to the training of AI models.[121]

Proposals for a global AI training data framework have gained increasing relevance towards ensuring that data is governed in the public interest.

6.2.2. Building public AI infrastructures

A second set of interventions concerns the infrastructures through which AI systems are developed and deployed. Current trajectories are marked by a high degree of concentration, with a small number of actors controlling access to data, compute, and model architectures. This concentration allows these actors to determine how knowledge is organized and accessed at scale, often aligning system design and deployment with commercial imperatives over broader public interest objectives. In contexts where access to compute, data, and technical capacity is limited, actors are constrained to using externally developed systems, reinforcing patterns of technological dependence.

These conditions point to the need to build public AI infrastructures as a countervailing force, as also pointed out in the Ibero-American and Asia-Pacific FGDs. While regulatory interventions can address specific harms, they do not alter the underlying concentration of infrastructural capacity.[122] Public AI infrastructure—encompassing compute resources, datasets, models, and institutional capacity—can provide the basis for more autonomous and context-responsive forms of technological development, equipping a wider range of actors to participate in AI innovation.

Public AI infrastructure—encompassing compute resources, datasets, models, and institutional capacity—can provide the basis for more autonomous and context-responsive forms of technological development, equipping a wider range of actors to participate in AI innovation.

119 European Union. (2024). Regulation (EU) 2024/... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Article 53: Obligations for providers of general-purpose AI models. EU Artificial Intelligence Act. https://artificialintelligenceact.eu/article/53/

120 United Nations. (2024). Governing AI for humanity: Final report of the High-level Advisory Body on Artificial Intelligence. https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf

121 UNESCO. (2025, September 23). Report of the Independent Expert Group on artificial intelligence and culture. https://www.unesco.org/sites/default/files/medias/fichiers/2025/09/CULTAI_Report%20of%20the%20Independent%20Expert%20Group%20on%20Artificial%20Intelligence%20and%20Culture%20%28final%20online%20version%29%201.pdf

122 Keller, P. (2026, forthcoming). Who Pays for the Commons? Three Conditions for an AI Levy That Works. In Knowledge Futures. Bot Populi https://botpopuli.net/section/knowledge-futures/

Emerging initiatives in this direction include publicly funded compute and model development programs, such as under the EuroHPC Joint Undertaking, which has begun to support AI model training through shared high-performance computing infrastructure.[123] Comparable efforts are also emerging in the Global South. The Brazilian Artificial Intelligence Plan (PBIA) includes significant public investments in compute capacity,[124] while the IndiaAI Mission has begun provisioning subsidized GPU infrastructure and supporting domestic actors developing foundational models.[125]

Public AI infrastructures can be designed to support the long-term stewardship of knowledge resources, including their preservation over time, and their use under defined and enforceable conditions. This includes the ability to differentiate between types of use, maintain accountability in downstream applications, and limit the unregulated appropriation of publicly generated data and knowledge.

The relevance of such arrangements is especially evident in domains where AI systems reorganize existing knowledge practices. Artisanal and traditional knowledge systems, such as those in small-scale fisheries or craft traditions, function as "living R&D [research and development] ecosystems," generating knowledge through embedded practice.[126] Yet they are often marginalized when their knowledge is incorporated into centralized AI models. Public infrastructures can enable these systems to interface with AI on more equitable terms, while retaining influence over how their knowledge is represented and used.

Paul Keller argues for the creation of a public AI ecosystem, in which AI-mediated access to information is not exclusively organized through commercial platforms. In such arrangements, institutions such as public service media, cultural heritage organizations, and open knowledge platforms can play a critical role in maintaining the context, provenance, and integrity of knowledge-functions that become even more important as AI systems increasingly mediate access to information.[127] Building such infrastructures requires sustained public investment alongside governance arrangements that support accountability and broad-based participation. Without these foundations, control over knowledge infrastructures is likely to remain concentrated, shaping not only how AI systems function, but who is able to shape them.

---

123 European High Performance Computing Joint Undertaking. (n.d.). EuroHPC Joint Undertaking. https://www.eurohpc-ju.europa.eu/index_en

124 Laboratório Nacional de Computação Científica. (2024, August 7). Plano Brasileiro de Inteligência Artificial (PBIA) 2024–2028. Governo do Brasil. https://www.gov.br/lncc/pt-br/assuntos/noticias/ultimas-noticias-1/plano-brasileiro-de-inteligencia-artificial-pbia-2024-2028 ; Brazil's Artificial Intelligence Plan (PBIA) of 2024: Enabler of AI sovereignty?, Johansson Neto, G. P., Farias da Costa, V. C., & Gaspar, W. B. (2024). Brazil's Artificial Intelligence Plan (PBIA) of 2024: Enabler of AI sovereignty? The African Journal of Information and Communication (AJIC), 34, 1–15. https://doi.org/10.23962/ajic.i34.20424

125 Press Information Bureau. (2025, October 12). Transforming India with AI: Over ₹10,300 crore investment & 38,000 GPUs powering inclusive innovation. PIB Delhi. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2178092; Vipra, J. (2025). Towards AI sovereignty: The good, the bad, and the ugly of AI policy in India. African Journal of Information and Communication, 35. https://doi.org/10.23962/ajic.i35.21263

126 Kurien, J. (2025, December 13). The hands that hold the future: Why artzanal knowledge is a living R&D ecosystem to learn from. Scroll.in. https://scroll.in/article/1088381/the-hands-that-hold-the-future-why-artizanal-knowledge-is-a-living-r-d-ecosystem-to-learn-from

127 Keller, P. (2025, June 30). Beyond AI and copyright: Funding a sustainable information ecosystem. Open Future.https://openfuture.eu/publication/beyond-ai-and-copyright/

6.2.3. Redistributing value

A third dimension concerns how value generated by AI systems is distributed. These systems derive economic returns from widely distributed cultural and informational resources, yet existing mechanisms—largely tied to copyright—capture only a limited subset of contributors. Efforts to design compensation systems within existing IP frameworks face practical constraints. As scholars have argued, the scale and non-deterministic nature of AI training make attribution difficult.[128] Compensation models based on tracing individual contributions are therefore hard to implement and tend to favor already visible and formalized forms of cultural production. Participants of the Africa FGD similarly questioned whether attribution-based compensation models can meaningfully operate in a context where AI outputs emerge through probabilistic aggregation rather than traceable acts of copying.

These constraints point to the need for redistributive mechanisms that operate at the level of systems, rather than individual works. As proposed by Keller, one approach is a levy on commercial AI services trained on publicly available data.[129] Applied at the stage of deployment of such services, such a levy shifts the focus from compensating discrete acts of appropriation to redistributing a portion of the economic value generated through large-scale data extraction and infrastructural control. Revenues collected in this way could be pooled and allocated to a range of public-interest institutions, including independent journalism and public service media, cultural and creative sector funds, open knowledge platforms, research and education institutions, and publicly governed data and AI infrastructure initiatives.

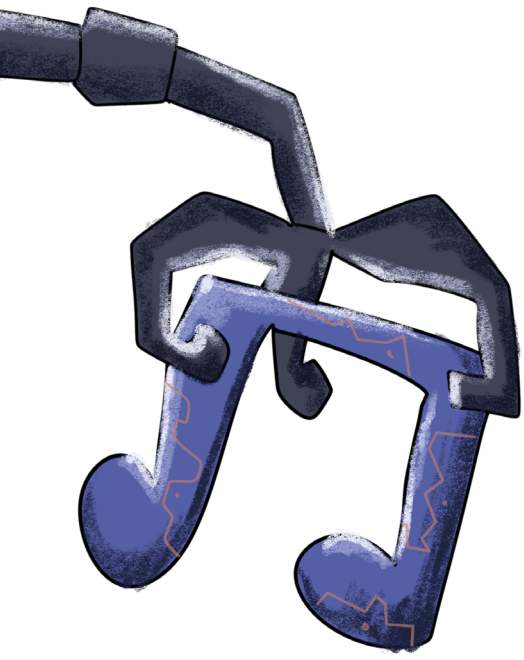
Importantly, this approach enables a potential linkage between value extraction and the financing of alternative AI ecosystems. Rather than attempting to trace individual contributions, revenues collected through such levies would be pooled and allocated to support the public infrastructures and community governance frameworks outlined earlier. Redistribution can thus be decoupled from attribution and reoriented toward rebuilding the material and institutional conditions of knowledge production.

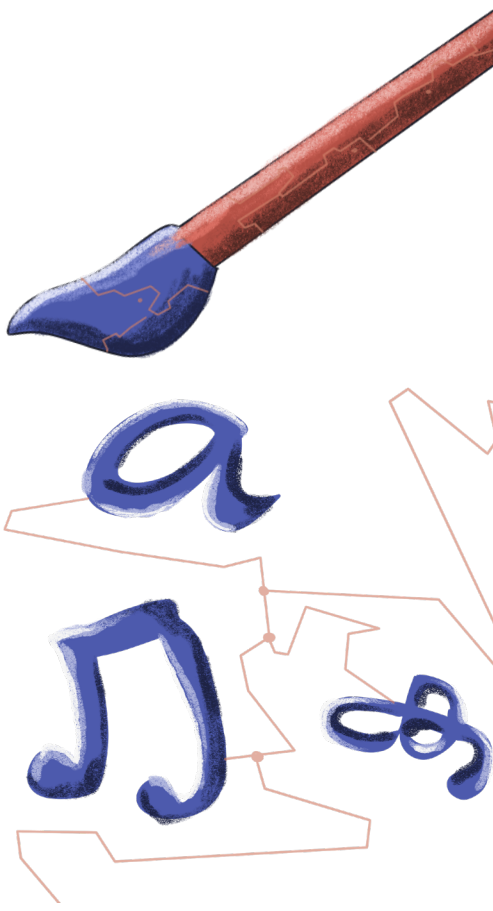
128 Lemley, M. A. (2021). Fair learning. Texas Law Review, 99(4), 743–784. https://texaslawreview.org/wp-content/uploads/2021/03/Lemley.Printer.pdf

129 Keller, P. (2026, forthcoming). Who Pays for the Commons? Three Conditions for an AI Levy That Works. In Knowledge Futures. Bot Populi https://botpopuli.net/section/knowledge-futures/; Keller, P. (2025, June 30). Beyond AI and copyright: Funding a sustainable information ecosystem. Open Future.https://openfuture.eu/publication/beyond-ai-and-copyright/

Rather than attempting to trace individual contributions, revenues collected through such levies would be pooled and allocated to support the public infrastructures and community governance frameworks outlined earlier. Redistribution can thus be decoupled from attribution and reoriented toward rebuilding the material and institutional conditions of knowledge production.

At the same time, such mechanisms raise significant questions around design and implementation, including how funds are governed, how beneficiaries are identified, and how such levies interact with existing regulatory and market structures. While not a complete solution, they indicate one possible pathway for addressing distributive imbalances and for reinvesting in the informational ecosystems on which AI systems depend.





Annex 1: List of participants of the Roundtable

1. Ahmed ElMoghazy, EiTESAL, Egypt (joined virtually)
2. Akshat Agrawal, AASA Chambers, India
3. Baarish A., Tattle, India
4. Debby Kristen, EngageMedia, Indonesia
5. Dhanaraj Thakur, Emerging Technology Initiative–Multiracial Democracy Project, George Washington University, United States of America
6. Hesham Dinana, EiTESAL, Egypt (joined virtually)
7. Jai Vipra, Cornell University, United States of America
8. Jim Thomas, Scan the Horizon, Canada (joined virtually)
9. KM Gopakumar, Third World Network, Malaysia
10. Laura Mantilla-León, Derechos Digitales, Chile
11. May Siksik, Innovation Network Global, Canada
12. Mila Samdub, AI Now Institute, United States of America
13. Nina Sangma, International Land Coalition, India
14. Pam Dixon, World Privacy Forum, United States of America
15. Paul Keller, Open Future Foundation, Netherlands (joined virtually)
16. Pooja Sood, Khoj International Artists' Association, India (joined virtually)
17. Ramya Chandrasekhar, French National Centre for Scientific Research, France (joined virtually)
18. Ranjit Singh, Data and Society, United States of America (joined virtually)
19. Rohini Lakshané, Independent Technologist and Researcher, India
20. Sriram Emani, IndianRaga, United States of America (joined virtually)
21. Sundar Athreya H., KIIT Law School, India
22. Swaraj Paul Barooah, SpicyIP, India
23. Valentine Goddard, AI Impact Alliance, Canada
24. Victor Pavarin, InternetLab, Brazil
25. IT for Change team: Abhineet Nayyar, Merrin Muhammed Ashraf, Nandini Chami, Saloni Mishra, Shobhit S

Annex 2: List of participants in the Regional Focus Group Discussions

Asia-Pacific Region

1. Adam Portelli, Media, Entertainment & Arts Alliance, Australia
2. Aditi Agarwal, Tech Journalist, India
3. Chandrabhanu Pattanayak, Institute of Knowledge Societies, India
4. Deborah Irene Christine, Tifa Foundation, Indonesia
5. Karen Lee Hizola, ETHIC, Philippines
6. Mahwash Fathima, Quantum Hub, India
7. Matt Byrne, Media, Entertainment & Arts Alliance, Australia
8. Nina Sangma, International Land Coalition, India
9. Padmini Ray Murray, Design Beku, India
10. Pyrou Chung, East-West Management Institute, Thailand
11. Revathi Sharma Kollegala, Center for International Forestry Research and World Agroforestry, India
12. Saai Sudharsan Sathiyamoorthy, Advocate, Madras High Court, India
13. Subhashish Panigrahi, O Foundation, Creative Commons India, India
14. Sundar Athreya H., KIIT Law School, India
15. Sushmita Viswanathan, Aapti Institute, India
16. Vinay Narayan, Aapti Institute, India

Africa Region

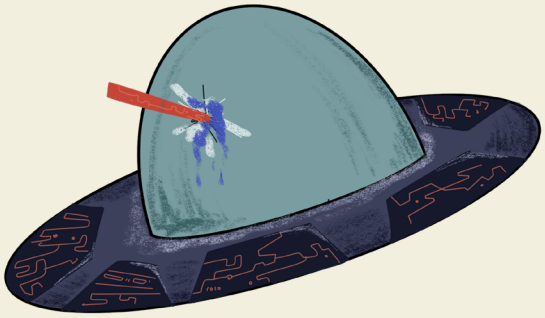
1. Andrew Rens, Research ICT Africa, South Africa
2. Cynthia Nzuki, Centre for Intellectual Property and Information Technology Law, Kenya
3. Dumisani Ndubane, Wikimedia South Africa, South Africa
4. Mardiya Siba Yahaya, Digital sociologist, South Africa
5. Nashilongo Gervasius, NUST Namibia, Namibia
6. Zanele, GRIT GBV, South Africa

Ibero-America Region

1. Alberto González Pulido OdeSIA, Spain
2. Carolina Botero, Fundación Karisma, Colombia
3. Gabriel Naranjo, TD SYNNEX Europe, Spain
4. Gabriela Busellini, Judiciary and UBA AI Lab, Argentina
5. Javier Chozas, Architect and academic; Director of a master's program at UNIR, Spain
6. Soledad Vogliano, ETC Group, Argentina

Middle East and Africa Region

1. Abdelmoneim Darwish, Injaz Consulting, Saudi Arabia
2. Adel Maged, Egyptian Court of Cassation, Egypt
3. Ahmed Darwish, Former Minister of Administrative Development, Egypt
4. Ahmed Elmaghraby, Vodafone Egypt, Egypt
5. Ahmed Elmoghazy, INTRO Technology Group, Egypt
6. Ahmed Saleh, African Union Advisory Group on AI & Ministry of Foreign Affairs, Egypt
7. Ahmed Tantawy, Applied Innovation Center, Egypt
8. Amira Saber, Egyptian Parliament, Egypt
9. Assem Wahby, Advanced Computer Technology, Egypt
10. Ebtehal Basiouny, Microsoft, Egypt
11. Engy Shehab, EFESO Management Consultants, Egypt
12. Hesham Dinana, EiTESAL, Egypt
13. Hossam Megahed, EiTESAL & Arab ICT Union (AICTU), Egypt
14. Jimson Olufuye, Kontemporary, Nigeria
15. Jotam Matariro, CAPASSO, South Africa
16. Maha Abd-nasser, Egyptian Parliament ICT Committee, Egypt
17. Mariam Fayez, Internet Masr, Egypt
18. Mohamed Salem, Former Minister of Communications and Information Technology, Egypt
19. Mohamed Shedeed, EiTESAL, Egypt
20. Omar Maher, Valify Solutions, Egypt
21. Samar Elsheikh, ITIDA, Egypt
22. Suzanne El Akabaoui, Personal Data Protection Center (PDPC), Egypt



IT for Change

https://itforchange.net/