Bridging the Workers' Data Value Gap in the Age of Corporate Automation

Mattias Rättzén*

TABLE OF CONTENTS

ABSTRACT

The rapid advancement of artificial intelligence ("AI") is transforming corporate innovation, enabling the automation of work traditionally performed by humans. Companies increasingly rely on internal data to develop and train AI systems, much of which originates from their own workforces. This includes work products created by employees or contractors in the course of their duties, here referred to as "workers' data." Workers' data may hold significant value as it is often of high quality, high quantity and of high relevance. Yet, ironically, if workers' data is used for automation purposes, it could displace the very employees who were responsible for generating the data in the first place. The risk of a "job apocalypse," where hundreds of millions of jobs could be replaced by AI in the coming years, becomes increasingly real as more companies push towards automating parts of their workforces.

Work products used to generate workers' data will often qualify as copyright-protected works. The workers, who are authors in copyright law, will often have assigned their copyright to their respective employers or contractees, whether by statute or by contract. However, that companies may own the copyright for the work products, including the data, does not necessarily confer unlimited freedom of use. This Article critically examines the legal boundaries on the use of workers' data for AI and automation purposes, focusing on the European Union (EU), the United Kingdom (UK), and the United States. Overall, the current legal framework largely falls short of adequately protecting workers' rights when it comes to use of their data by employers, with one important exception. In the EU, the DSM Directive introduces a right for authors to claim additional remuneration where the economic value derived from their works is significantly greater than what they were originally paid. This right to contract adjustment, commonly referred to as the "best-seller" rule, may, in certain cases, help workers whose data are being used without fair remuneration, bridging what is coined as the "workers' data value gap." Whether workers are entitled to further remuneration will depend on highly fact-specific circumstances, and outcomes are likely to vary from case to case.

The urgency of addressing these issues cannot be overstated. Workers' data is expected to play an increasingly central role in corporate automation projects worldwide, with consequences that extend far beyond the EU. Without a clear and coherent legal framework, there is a dual risk. On the one hand, companies may undervalue or fail to properly compensate for the human contributions that make automation possible; on the other hand, legal uncertainty or excessive compensation claims could deter or slow down automation initiatives. This Article calls for economic, policy, and legal research to explore these questions in greater detail, and to develop balanced solutions that both protect workers' rights and foster innovation in the coming age of corporate automation.

INTRODUCTION

Imagine that you have been working for a company for a decade. Over the years, you have dedicated countless hours to creating reports, strategies, media, and other creative or proprietary content. All of this work, as per the terms of your contract for employment or independent contractorship, or by default under applicable law, belongs to the company. One day, your company decides to take all this work and feed it as data into a generative artificial intelligence ("AI") model or system. The model or system will be trained on the unique characteristics of your contributions, making it capable of replicating, or even surpassing, your own work. Soon after, you are handed your termination notice. AI has replaced you.

This hypothetical scenario is increasingly becoming a reality for many workers. For example, IBM recently made plans to replace up to 30% of its back-office roles with AI over the next five years, potentially impacting close to 7,800 jobs.[1] Similarly, BT Group aims to cut up to 55,000 jobs by 2030, approximately 40% of its workforce, with 10,000 of those positions expected to be replaced by AI.[2] Other companies have already started their AI transition journey. In 2023, the Indian e-commerce company Dukaan replaced 90% of its customer support staff with an AI chatbot.[3] The Institute for Public Policy Research has estimated that up to 8 million jobs would be lost in the UK alone due to AI, without government intervention.[4] Another report

---

* LL.M. (Harvard), PGDip (Oxford), LL.M. (Lund). Mattias Rättzén is an intellectual property and artificial intelligence attorney admitted in New York, England and Wales, and Sweden, and is the author of multiple published articles. In addition to working in private practice in London and Stockholm, he has been a member of the Swedish SCCR delegation at WIPO in Geneva. The views in this Article are the author's own.

1. Brody Ford, *IBM's CEO Expects A.I. to Be so Good at Back Office Work that He Plans to Pause Hiring Humans for Those Jobs*, FORTUNE (May 1, 2023), https://fortune.com/2023/05/01/ibm-ceo-ai-artificial-intelligence-back-office-jobs-pause-hiring/ [https://web.archive.org/web/20250918150440/https://fortune.com/2023/05/01/ibm-ceo-ai-artificial-intelligence-back-office-jobs-pause-hiring/].

2. Mark Sweney, *BT to Axe Up to 55,000 Jobs by 2030 as it Pushes into AI*, GUARDIAN (May 13, 2023), https://www.theguardian.com/business/2023/may/18/bt-cut-jobs-telecoms-group-workforce [https://web.archive.org/web/20250918150959/https://www.theguardian.com/business/2023/may/18/bt-cut-jobs-telecoms-group-workforce].

3. Anna Cooban, *This CEO Replaced 90% of Support Staff with an AI Chatbot*, CNN (July 12, 2023), https://edition.cnn.com/2023/07/12/business/dukaan-ceo-layoffs-ai-chatbot/index.html [https://web.archive.org/web/20250918151650/https://edition.cnn.com/2023/07/12/business/dukaan-ceo-layoffs-ai-chatbot/index.html].

4. Carsten Jung & Bhargav Srinivasa Desikan, *Transformed by AI: How Generative Artificial Intelligence Could Affect Work in the UK—and How to Manage it*, INST. PUB. POL'Y RSCH. (Mar. 2024), https://ippr-org.files.svdcdn.com/production/Downloads/Transformed_by_AI_March24_2024-03-27-121003_kxis.pdf

from Goldman Sachs has estimated that AI could replace the equivalent of 300 million full-time jobs worldwide.[5] The World Economic Forum has forecasted the same, surveying that a staggering 41% of employers worldwide intend to downsize their workforces by 2030 due to AI.[6] Employees themselves have also admitted that up to 31% of their work tasks are replaceable by AI, project management and collaboration software.[7]

This job apocalypse, although devastating for many workers, will result in significant efficiency gains and cost savings for companies, which have a legitimate interest in improving and streamlining their operations. There is currently no general prohibition against using AI to replace workers, which merely reflects the continuation of patterns seen in labor markets for many decades. Historically, advancements in new technologies have led to the displacement of human workers, particularly in roles which have been repetitive in nature or of less value, but have also created new opportunities for those displaced. But what makes the present situation different from the past, however, is that workers' data is being used against them. Data is the new gold in the fourth industrial revolution.[8] AI systems and models often require vast amounts of data to learn and perform their tasks, and it is the quality, quantity, and diversity of that data which determine their performance and accuracy.[9] The data produced by workers, originating from their work products, will be especially valuable to companies as it is often of high

---

[https://web.archive.org/web/20250918151907/https://ippr-org.files.svdcdn.com/production/Downloads/Transformed_by_AI_March24_2024-03-27-121003_kxis.pdf].

5. GOLDMAN SACHS, *Generative AI Could Raise Global GDP by 7%*, ARTIFICIAL INTELLIGENCE (Apr. 5, 2023), https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent [https://web.archive.org/web/20250918152304/https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent].

6. WORLD ECON. FORUM, INSIGHT REPORT: FUTURE OF JOBS REPORT 63 (Jan. 2025), https://reports.weforum.org/docs/WEF_Future_of_Jobs_Report_2025.pdf [https://web.archive.org/web/20251003183637/https://reports.weforum.org/docs/WEF_Future_of_Jobs_Report_2025.pdf].

7. ASANA, The STATE OF AI AT WORK 7 (2024), https://assets.asana.biz/m/25388d9dbeadd06e/original/FY25_Q2_State-of-AI-at-Work-Anthropic.pdf [https://web.archive.org/web/20250918152926/https://assets.asana.biz/m/25388d9dbeadd06e/original/FY25_Q2_State-of-AI-at-Work-Anthropic.pdf] (a survey consisting of more than 5,000 knowledge workers in the UK and the United States).

8. *See, e.g.*, Ding Wang, Shantanu Prabhat & Nithya Sambasivan, *Whose AI Dream? In Search of the Aspiration in Data Annotation*, ARXIV 1 (Mar. 21, 2022) ("Data is fundamental to AI/ML Models."); Laura Galindo, Karine Perset & Francesca Sheeka, *OECD Going Digital Toolkit: An Overview of National AI Strategies and Policies*, OECD 10–11 (2021), www.oecd.org/content/dam/oecd/en/publications/reports/2021/08/an-overview-of-national-ai-strategies-and-policies_913b6e4b/c05140d9-en.pdf [https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/08/an-overview-of-national-ai-strategies-and-policies_913b6e4b/c05140d9-en.pdf] ("Data access and sharing are key to accelerating AI uptake.").

9. For a more detailed description of how generative AI technologies use training data, *see* Mattias Rättzén, *Location Is All You Need: Copyright Extraterritoriality and Where to Train Your AI*, 26 COLUM. SCI. & TECH. L. REV. 175, 182–91 (2024).

relevance to the organization and of high quality. In the years to come, workers' data could become one of the most valuable assets for many companies as they are rolling out more AI systems, some of which may well be used to make certain human tasks redundant. Yet workers are so far not realizing any of this increased economic value, which suggests that there could potentially be, in some circumstances, a so-called "value gap" for workers' data.

The use of workers' data for corporate automation purposes raises complex legal questions that require urgent attention. This Article critically discusses to what extent companies may lawfully use workers' data and the consequences of doing so. Focusing on the legal situation in the EU, UK, and the United States, it explores whether there are any restrictions for such use, notwithstanding that the copyright for any work products, and any data contained therein, has been assigned to the company by statute or contract. The most important of these restrictions is the so-called "best-seller" rule in the EU DSM Directive, which entitles authors, including workers, to adjust contracts previously entered into for the exploitation of their copyrighted works.[10] The trigger in these circumstances will be if there is a significant discrepancy between the received and realized economic value of those works. This Article discusses to what extent there is a workers' data value gap, which has not been bridged, and whether workers could claim further remuneration for the exploitation of their works as data. The need to address these questions and provide clarity could not be more pressing, as workers' data is expected to be used to a considerable extent in the coming years for corporate automation purposes, potentially resulting in many millions losing their jobs. The impact of this discussion will extend far beyond workers and companies based in the EU, but also elsewhere in the world. A balanced approach will be essential in this regard to ensure that both workers and companies benefit fairly from the economic value derived from workers' data in the age of corporate automation.

I. DEFINING WORKERS' DATA

Whether as employees or independent contractors, the role of many workers is to produce various forms of content, manage and organize projects, or to communicate matters, resulting in proprietary work products and information.[11] This output can be classified into several categories, by way of example:

1. documents: written materials, in any format;
2. designs: graphic designs, schematics, and other visual or structural plans;
3. code: computer programs, scripts, and algorithms;

10. Directive 2019/790, of the European Parliament and of the Council of 17 April 2009 on Copyright and Related Rights in the Digital Single Market and Amending Directives 96/9/EC and 2001/29/EC, 2019 O.J. (L 130) 92 [hereinafter, the "DSM Directive"].

11. The term "workers" in this Article is used to encompass both employees and independent contractors.

4. databases: documents or information collected and arranged in a systematic or methodical way;
5. marketing: brochures, website and social media content, or advertisements;
6. spreadsheets: financial forecasts, market analyses, or similar content for other purposes;
7. communications: emails, letters, instant messages, calls, and other forms of workplace communication, whether internal or external; and
8. media: music, audiovisuals, recordings, and photographs.

These tangible work products will be created by workers in the course of carrying out their regular work duties. The output is typically the direct result of the worker's intellectual or creative efforts and is often subject to copyright or related rights protection as literary, dramatic, musical, or artistic works, or as recordings, broadcasts, and databases. Although the work products as such will have tangible value, companies are increasingly recognizing work products as important data assets, which are of high quality, relevant, and which are exclusive to the organization. All work products will be composed of some form of data, or here referred to as "workers' data," which can be sourced, organized, analyzed, and used for various purposes, including corporate automation.

II. HOW COMPANIES ARE USING WORKERS' DATA FOR AUTOMATION PURPOSES

Companies are increasingly leveraging the vast amounts of data produced by their workers to drive corporate automation efforts. This trend is particularly evident in industries where repetitive workers' tasks can be easily automated through AI and machine learning, such as administration and project management, but also for areas such as marketing, sales, and customer service.

Companies have so far been adopting a mix of third party and in-house solutions when it comes to AI deployment. According to an executive survey conducted in 2023 by MIT Sloan Management Review and Boston Consulting Group, consisting of more than 1,240 respondents representing companies in fifty-nine industries and eighty-seven countries, 78% of organizations use third party AI tools, a figure that is likely to be even higher today.[12] More than half of the responding organizations used third party tools exclusively, having no internally designed or developed AI technologies.[13] Yet an increasing number of companies that have the resources to

12. Elizabeth M. Renieris, David Kiron & Steven Mills, *Building Robust RAI Programs as Third-Party AI Tools Proliferate*, MIT SLOANE MGMT. R. 1, 4 (2023), https://sloanreview.mit.edu/projects/building-robust-rai-programs-as-third-party-ai-tools-proliferate/ [https://web.archive.org/web/20250927022926/https://web-assets.bcg.com/1b/18/c684f0174e088e068efc4c62c942/building-robust-rai-programs-as-third-party-ai-tools-proliferate.pdf].

13. *Id.* at 4.

build their own AI tools are also doing so. For example, Bank of America informed investors on an earnings call in 2023 that they are using internal AI tools to leverage a proprietary predictive language program tailored specifically to the bank's unique data sets.[14] Meta has also rolled out their own chatbot, called Metamate, to its employees, which was trained on internal data and uses such data to assist employees in summarizing meetings, writing code, and debugging features.[15] PwC has similarly developed its own internal chatbot.[16] While many of these tools are so far being used to improve business operations, companies are increasingly also using AI to replace human tasks. A study from October 2022 estimated that 14% of workers had, at that time, experienced job displacement due to automation or AI.[17] Another report estimated that, from more than 80,000 job losses in the United States in May 2023, approximately 5% could be directly attributed to AI.[18] For example, the Swedish fintech company Klarna has reduced its workforce from 5,000 to 3,800 in the past year, partly due to AI.[19] Klarna recently stated that "[o]ur AI assistant now performs the work of 700 employees, reducing average resolution time from eleven minutes to just two, while maintaining the same customer satisfaction scores as human agents."[20] Meanwhile, Klarna managed to increase its annual revenue by 27% to $1.3 billion.[21]

How companies use worker-produced content as a data source for AI depends on context. Many now integrate third party AI solutions into their businesses, but this does not mean they rely solely on off-the-shelf data. For example, fine-tuning

14. *Companies Tap Their Own Data to Drive Efficiencies with AI*, PYMNTS (Apr. 26, 2023), https://www.pymnts.com/news/artificial-intelligence/2023/companies-tap-their-own-data-drive-efficiencies-with-ai/ [https://perma.cc/ATZ5-LVFA].

15. ABP NEWS BUREAU, *Meta's AI Chatbot Is Here and It's Called Metamate*, ABP: TECHNOLOGY (June 13, 2023), https://news.abplive.com/technology/meta-ai-chatbot-metamate-launch-mark-zuckerberg-internal-data-verge-1608614 [https://web.archive.org/web/20251003194037/https://news.abplive.com/technology/meta-ai-chatbot-metamate-launch-mark-zuckerberg-internal-data-verge-1608614].

16. Tess Bennett, *Meet ChatPwC, the Custom-Built AI Tool Being Rolled Out at the Firm*, FIN. R. (Feb. 6, 2024), https://www.afr.com/companies/professional-services/meet-chatpwc-the-custom-built-ai-tool-being-rolled-out-at-the-firm-20240202-p5f21q [https://web.archive.org/web/20251003194818/https://www.afr.com/companies/professional-services/meet-chatpwc-the-custom-built-ai-tool-being-rolled-out-at-the-firm-20240202-p5f21q].

17. Eric Dahlin, *Are Robots Really Stealing Our Jobs? Perception Versus Experience*, 8 SOCIUS (2022).

18. *Challenger Report: Layoffs Jump on Tech, Retail, Auto; YTD Hiring Lowest Since 2016*, CHALLENGER, GRAY & CHRISTMAS (Jun. 1, 2023), https://omscgcinc.wpenginepowered.com/wp-content/uploads/2023/06/The-Challenger-Report-May23.pdf.

19. Tom Gerken, *Klarna: AI Lets Us Cut Thousands of Jobs—But Pay More*, BBC (Aug. 28, 2024), https://www.bbc.co.uk/news/articles/c80e1gp9m9zo [https://web.archive.org/web/20251105171308/https://www.bbc.com/news/articles/c80e1gp9m9zo].

20. Press Release, Klarna, Klarna AI Assistant Handles Two-thirds of Customer Service Chats in Its First Month (Feb. 27, 2024), https://www.klarna.com/international/press/klarna-ai-assistant-handles-two-thirds-of-customer-service-chats-in-its-first-month [https://web.archive.org/web/20251105172044/https://www.klarna.com/international/press/klarna-ai-assistant-handles-two-thirds-of-customer-service-chats-in-its-first-month/].

21. Gerken, *supra* note 19.

often requires specialized training data to improve performance in a given field.[22] The prospect for many AI systems lies in the data, and internal company data will be particularly valuable.[23] A recent Harvard Business Review article explained that turning third party AI into competitive advantage ultimately depends on powering it with unique, internal data.[24] Such data is highly relevant, high-quality, and produced by experienced staff. Because no other firm has access to it, the data is exclusive and, when accumulated over time, also abundant. Companies are therefore expected to increasingly treat workers' data as a core corporate asset for AI development and deployment.

There is limited public disclosure of how companies use workers' data in AI systems, or how this affects employment, including whether it prompts redundancies or restructuring. Yet lack of transparency does not mean it is not occurring. For example, a graphic designer who lost his job stated in a YouTube video in May 2024, that "[b]asically, all of the material that I have provided over the past six years is now being fed to AI and templated. [. . .] A design that takes me thirty minutes now takes the AI thirty seconds, as it has been trained on all my templates. Essentially, I think it literally reuses my templates."[25] After being made redundant, he found competitors doing the same, some even advertising it. One example is not conclusive, but similar accounts abound.[26] Workers' data is increasingly used to train AI systems, sometimes making the very workers who produced it redundant.

III. WHO OWNS THE COPYRIGHT OF WORKERS' DATA?

A fundamental question is who owns the data produced by workers, whether employees or independent contractors. Because data is, as discussed above, a byproduct of the work products themselves, the ownership of the latter determines the former. Work products are often protected by copyright as forms of literary, dramatic, musical, or artistic works. Databases, which are systematic collections of independent works, data, or other materials, may also be protected by copyright or

22. Jiawei Zheng et al., *Fine-tuning Large Language Models for Domain-specific Machine Translation*, ARXIV 1, 2 (Feb. 23, 2024), https://arxiv.org/abs/2402.15061v1 [https://web.archive.org/web/20250918155658/https://arxiv.org/abs/2402.15061v1].

23. *See* Wang, Prabhat & Sambasivan, *supra* note 8.

24. Scott Cook, Andrei Hagiu & Julian Wright, *Turn Generative AI from an Existential Threat into a Competitive Advantage*, HARV. BUS. R. (Jan. 2024), https://hbr.org/2024/01/turn-generative-ai-from-an-existential-threat-into-a-competitive-advantage [https://web.archive.org/web/20251105172941/https://hbr.org/2024/01/turn-generative-ai-from-an-existential-threat-into-a-competitive-advantage].

25. NADESTRAIGHT, *I Lost My Job This Week Because of AI*, at 2:16 (YouTube, May 3, 2024) https://www.youtube.com/watch?v=U2vq9LUbDGs [https://web.archive.org/web/20250927023929/https://www.youtube.com/watch?v=U2vq9LUbDGs].

26. *See, e.g.*, Mark Pesce, *AI Stole My Job and My Work, and The Boss Didn't Know—or Care*, REGISTER (Aug. 15, 2024), https://www.theregister.com/2024/08/15/robot_took_my_job/ [https://web.archive.org/web/20250918182832/https://www.theregister.com/2024/08/15/robot_took_my_job/].

separately as a related or *sui generis* right. There is, therefore, no such thing as copyright ownership over "data" as such, but only protected works which may comprise data.[27] Employment and contractor agreements typically allocate copyright ownership in any work products. If the contract is silent on the issue, ownership may arise as an implied term, or copyright laws will set out what is the default statutory rule.

Copyright laws tend to favor employers in this regard. In the UK and the United States, works created by employees in the course of employment belong to the employer under the "work-for-hire" doctrine, unless parties agree otherwise.[28] In the United States, contractees may also own works commissioned from contractors if they fall into certain narrowly defined statutory categories, such as collective works, films, and compilations.[29] In the UK, contractors remain default owners unless the contract assigns the copyright.[30] What complicates this picture is that copyright is a territorial right. There is no such thing as international copyright, which means that, where an employee or contractor produces a work, there will be a bundle of territorially distinct rights in each country where the work is capable of being protected.[31] Therefore, even if an employee creates a work when residing in the United States or the UK, where copyright will transfer automatically to the employer by way of statute, the employee will simultaneously and independently be afforded copyright under any foreign statutes, unless the worldwide copyright portfolio is contractually assigned.

Copyright ownership in the course of employment is not harmonized in the EU, and differs substantially compared to the Anglo-American approach.[32] Continental European copyright law, although the subject of each Member State, generally does

27. *See* Lothar Determann, *No One Owns Data*, 70 HASTINGS L.J. 1, 18–19 (2019) (discussing that copyright laws do not provide protection to the underlying data itself). Similarly, the U.S. Supreme Court explained in *Feist Publications, Inc. v. Rural Telephone Service Co.* that, "raw data does not satisfy the originality requirement." 499 U.S. 340, 361 (1991).

28. Copyright, Designs and Patents Act 1988, c. 48, §11 (UK) ("Where a literary, dramatic, musical or artistic work is made by an employee in the course of his employment, his employer is the first owner of any copyright in the work subject to any agreement to the contrary."); 17 U.S.C. § 201(b) ("In the case of a work made for hire, the employer or other person for whom the work was prepared is considered the author for purposes of this title, and, unless the parties have expressly agreed otherwise in a written instrument signed by them, owns all of the rights comprised in the copyright.").

29. 17 U.S.C. § 101 (defining "work made for hire").

30. *See* Griggs Group Ltd & Others v. Evans & Others [2003] EWHC (Ch) 2914 (UK).

31. *See, e.g.*, Jane C. Ginsburg, *International Copyright: From a "Bundle" of National Copyright Laws to a Supranational Code?*, 47 J. COPYRIGHT SOC'Y U.S.A. 265, 266 (2000). This fundamental principle of national interdependence is confirmed in Article 5 of the Berne Convention for the Protection of Literary and Artistic Works. *See* Berne Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886, as revised at Paris, July 24, 1971, and as amended on Sept. 28, 1979, S. Treaty Doc. No. 99–27, 1161 U.N.T.S. 3 [hereinafter, the "Berne Convention"]. Specifically, Article 5(2) provides that authors shall enjoy and exercise their rights in all signatory states, independently of the existence of protection in the country of origin of the work. *Id.* at art. 5(2).

32. Dénes Legeza, *Employer as Copyright Owner from a European Perspective*, SERCI ANNUAL CONGRESS 3 (Working Paper, Sept. 2015), http://www.serci.org/congress_documents/2015/Legeza.pdf [https://web.archive.org/web/20250927060411/http://www.serci.org/congress_documents/2015/Legeza.pdf] (discussing the interface between employment and copyright in Europe).

not recognize a work-for-hire doctrine, meaning that copyright will originally belong to the author who initially created the work.[33] In Germany, for example, employers do not acquire copyright even when works are created in the course of employment, though employees may grant exploitation rights to fulfill contractual duties.[34] Similarly, in France[35], Belgium[36] and Sweden[37] the default rule is that the employee, as the author, is presumed to be the copyright owner, unless assigned. The situation will be the opposite where assignment is an express or implied term in the employment contract, which will ultimately be a question of interpreting the contract.[38] The position will be the same for independent contractors as for employees. In other European countries, like Spain, the Netherlands, Poland, and Hungary, the default beneficiary of copyrighted works in employment relationships will be the employer.[39]

Unlike other types of works, authorship of computer programs and rights exercised in respect of such works is harmonized within the EU as part of the Computer Programs Directive 2009/24/EC.[40] Computer programs are defined as programs in any form, including preparatory design work.[41] While authorship vests in the natural person(s) "creating" the program, Article 2(3) of the Directive provides that, where created by an employee in the course of employment, the employer exclusively enjoys the economic rights, unless the contract provides otherwise.[42] Though framed as economic rights rather than ownership, employers

33. Legeza, *supra* note 32, at 23 (concluding that there is not a work-for-hire copyright doctrine in most European countries, with the exception of the Netherlands).

34. *See* PREU BOHLIG, IP ASSIGNMENT CLAUSES IN GERMAN EMPLOYMENT CONTRACTS (2019), https://preubohlig.de/wp-content/uploads/2019/07/IP-Assignment-Clauses-in-International-Employment-Contracts.pdf [https://web.archive.org/web/20250919130708/https://preubohlig.de/wp-content/uploads/2019/07/IP-Assignment-Clauses-in-International-Employment-Contracts.pdf].

35. *See* Jean-Luc Piotraut, *An Authors' Rights-Based Copyright Law: The Fairness and Morality of French and American Law Compared*, 24 CARDOZO ARTS & ENT. L.J. 549, 577 (2006).

36. *See* SÉVERINE DUSOLLIER ET AL. (EUR. PARLIAMENT DIRECTORATE-GENERAL FOR INTERNAL POLICIES), CONTRACTUAL ARRANGEMENTS APPLICABLE TO CREATORS: LAW AND PRACTICE OF SELECTED MEMBER STATES 44 (2014).

37. *See* MATS GLAVÅ, ARBETSRÄTT 435 (4th ed. 2020).

38. *See* Piotraut, *supra* note 35 at 577 n.214; Upphovsrätten på den digitala inre marknaden, prop. 2021/22:278 [Copyright in the Single Digital Market Bill] (Swed.) [hereinafter "prop. 2021/22:278"].

39. *See* DUSOLLIER ET AL., *supra* note 36, at 44–45. Regarding the Netherlands, see Auteurswet van 1912 [Copyright Act of 1912] art. 7 (Neth.), as amended Stb. 2021, 248. For an unofficial English translation, see THE DUTCH COPYRIGHT ACT 2021–2022: AUTEURSWET 2021–2022 (Visser, Schaap & Kreijger eds., Hendriks & James Legal Translations trans., 2021) (ebook) [hereinafter "Dutch Copyright Act"]. Recently, the European Court of Justice ruled in *Orchestre national de Belgique* that national legislation that automatically assigns performers' rights to their employers, without prior consent from those performers, is not permitted under Directive 2001/29/EC. *See* Case C-575/23, FT v. Belgian State, ECLI:EU:C:2025:141, ¶ 128 (Mar. 6, 2025). Although that case concerned performers' rights, it is possible that a similar position may also be taken for copyright.

40. Directive 2009/24/EC, of the European Parliament and of the Council of 23 April 2009 on the Legal Protection of Computer Programs, 2009 O.J. (L 111) 16.

41. *Id.* at recital 7.

42. *Id.* at art. 2(1), art. 2(3).

are effectively granted full control over any reproduction, adaptation, translation, and exploitation.[43]

In practice, default statutory copyright rules matter little when contracts address ownership. Employment and contractor agreements routinely assign copyright to the employer or contractee, and has become "boilerplate" in many industries. [44] In such cases, employees or contractors generally cannot claim further compensation for later use of their works. For companies, assigning copyright ownership is in fact important to protect themselves against such claims.

Not all work products will attract copyright. Protection extends only to original expressions fixed in a tangible medium,[45] and mere facts, data, and abstract ideas fall outside its scope.[46] Blank templates and forms that merely record information are similarly unprotected. However, contracts can, and frequently do, govern to what extent the employer or contractee is afforded *commercial ownership*, or may exercise any contractual rights, over any work product, in any form, whether or not such work products are protected by copyright.[47] The freedom of contract will trump in this regard in the absence of any statutory rights to the contrary.

IV. RESTRICTIONS ON USING WORKERS' DATA

A. THE PERSONAL DATA CLAWBACK

Data privacy laws regulate the processing of personal data. Personal data is broadly defined as any information relating to an identified or identifiable natural person, also known as a "data subject."[48] This includes information enabling direct

43. *Id.* at art. 4(1).

44. Elizabeth Knuppel, *"A Mortgage on a Man's Brain": The Unconscionability of Overly Broad Intellectual Property Assignment Clauses in Employment Contracts*, 100 TEX. L. REV. 971 (2022) (discussing the pervasiveness of intellectual property assignments in employment contracts and to what extent these could be considered unconscionable).

45. 17 U.S.C. § 102(a); Case C-145/10, Eva-Maria Painer v. Standard VerlagsGmbH, ECLI:EU:C:2011:798, ¶ 89 (Dec. 1, 2011); Case C-604/10, Football Dataco Ltd v. Yahoo! UK Ltd, ECLI:EU:C:2012:115, ¶ 38 (Mar. 1, 2012) (holding that, with respect to databases, the "criterion of originality is satisfied when, through the selection or arrangement of the data which it contains, its author expresses his creative ability in an original manner by making free and creative choices . . . and thus stamps his 'personal touch'").

46. *See, e.g.*, Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991).

47. The significance of contracts for protecting data, in the absence of copyright protection, has been highlighted by the European Commission and in European studies on several occasions. *See, e.g.*, *Building a European Data Economy*, at 10, COM (2017) 9 final (Jan. 10, 2017) (considering that contracts "might be a sufficient response" to address the absence of statutory protection for data); *Study on Model Contract Terms and Fairness Control in Data Sharing and in Cloud Contracts and on Data Access Rights: Final Report*, EUR. COMM'N (Apr. 2022), https://op.europa.eu/publication-detail/-/publication/dfb3a486-e6d4-11ec-a534-01aa75ed71a1
[https://web.archive.org/web/20251010113556/https://op.europa.eu/en/publication-detail/-/publication/dfb3a486-e6d4-11ec-a534-01aa75ed71a1].

48. Regulation (EU) 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free

or indirect identification, such as names, location data, online identifiers, or attributes tied to physical, physiological, genetic, mental, economic, cultural, or social identity.[49] Work products such as documents and correspondence will frequently contain workers' personal data.

Data privacy laws do not create ownership rights over personal data, but grant data subjects certain control rights resembling ownership.[50] Processing of personal data is only permissible if there is a lawful basis. In the EU, the GDPR sets out, among others, the data subject's consent and legitimate interests as two such bases.[51] Recital 47 further explains that a legitimate interest may exist where there is a relevant relationship between controller and data subject, such as employment.[52] Employers and contractees may claim they have a legitimate interest in processing workers' data in any produced works, though whether this extends to processing for automation is less clear. In practice, employment and contractor agreements routinely secure consent for broad processing purposes. If that includes automation, then employers or contractees will be in a position to rely on such consent, if it is freely given. Employers and contractees may also rely on the necessity of processing for performance of the contract where workers' personal data is embedded in work products, though this will depend on the particular circumstances.[53]

The GDPR famously extends its scope extraterritorially, which has led to the globalization of European data privacy standards.[54] It applies to the processing of data subjects in the Union, regardless of the controller's or processor's location, where processing relates to the offering of goods or services to EU subjects or to activities of an EU establishment.[55] Accordingly, it applies both to workers in the EU and to controllers or processors established there.

The GDPR also grants a right to erasure, or "right to be forgotten," which is triggered at least when consent is withdrawn or there is no other lawful ground for the processing.[56] Consent may be freely withdrawn at any time, which is a statutory

---

Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1, at art. 4(1) [hereinafter, the "GDPR"].

49. *Id*.

50. Lothar Determann, *No One Owns Data*, 70 HASTINGS L.J. 1, 26-27 (2018).

51. GDPR, *supra* note 48, at art. 6(1). Unlike the GDPR, the California Consumer Privacy Act [hereinafter, the "CCPA"], which is the most far-reaching data privacy law enacted in the United States, mainly focuses on effectuating data privacy rights by requiring companies to inform consumers when their personal data is collected, as opposed to only permitting such processing where there is a legal basis. *See* Cal. Civ. Code § 1798.100–1798.199.100 (2018). The CCPA also applies to "consumers," which is defined to include California-based employees.

52. GDPR, *supra* note 48, at recital 47.

53. *Id.* at art. 6(1)(b).

54. For a more detailed discussion on the globalization of the GDPR, *see* Simon Gunst & Ferdi De Ville, *The Brussels Effect: How the GDPR Conquered Silicon Valley*, 26 EUR. FOREIGN AFF. REV. 437 (2021).

55. GDPR, *supra* note 48, at art. 3(1)–(2).

56. *Id.* at art. 17(1)(b). Another triggering ground for the right to be forgotten is where the data subject objects to the processing, however this is only available where there are no overriding legitimate grounds for the processing. *See* art. 17(1)(c), which refers to art. 21.

right that cannot be waived by contract.[57] If consent is waived, the lawfulness of any prior processing remains unaffected.[58] Employers' or contractees' fallback is to identify another legal ground for processing, or to anonymize the personal data, such that it no longer qualifies as personal data.[59] Anonymization is increasingly used to address privacy concerns, including in AI contexts, but data privacy laws such as the GDPR may still apply if re-identification remains possible, in which case the anonymization was never fully effective to begin with.[60]

B. CONTRACT LAW RESTRICTIONS ON USING WORKERS' DATA

The freedom of contract is strong, and there are few statutory restrictions for assigning copyright or processing workers' data. However, contract law does impose certain limitations, including that contracts or specific clauses can be rendered unenforceable on the basis of unconscionability or undue influence.

A bargain cannot be unfair and unconscionable, and courts in the UK and the United States have long recognized that even voluntary agreements may be voidable if that is the case. In *Lloyds Bank v. Bundy*, Lord Denning described unconscionability under the rubric of inequality of bargaining power, protecting those who, without independent advice, enter into contracts on grossly unfair terms while impaired by poverty, ignorance, or infirmity, and subject to undue pressure.[61] English law has come to distill this doctrine into three separate requirements: (i) one party has been at a serious disadvantage to the other, (ii) the other party has exploited this disadvantage in a morally culpable manner, and (iii) the resulting transaction has been overreaching and oppressive.[62] The first requirement, serious disadvantage, can include poverty, ignorance, and lack of independent advice, but it is debated whether a contract should be voidable if one party was pressured by circumstances entirely outside the other party's influence.[63] The second requirement, exploitation, demands a "high level of knowledge" of the disadvantage.[64] The third requirement, focused on the outcome, assumes that the contractual terms were imposed in a morally reprehensible manner, and the other party knowingly exploited the vulnerability.[65] These are difficult conditions for workers to meet in the context of computational use of data, especially where

57. *Id.* at art. 7(3)–(4) and recital 42.

58. *Id.* at art. 7(3).

59. *Id* at recital 26.

60. *Id.* (defining anonymous data as instances where "the data subject is not or no longer identifiable"); *see also* Alvaro Moretón & Ariadna Jaramillo, *Anonymisation and Re-Identification Risk for Voice Data*, 7 EUR. DATA PROT. L. REV. 274 (2021) (discussing the risks of re-identification, in the case of voice data, using measures such as inference or linkage).

61. Lloyds Bank v. Bundy [1975] QB 326 at 339 (UK).

62. Alec Lobb (Garages) Ltd. v. Total Oil (G.B.) Ltd. (1984) 1 WLR 87 at 94–95 (UK).

63. *Id.*

64. *See* Rick Bigwood, *Contracts by Unfair Advantage: From Exploitation to Transactional Neglect*, 25 OXF. J. LEG. STUD. 65, 70–71 (2005).

65. Mitchell v. James [2001] All ER (Ch) at 116 (UK).

contractual terms were reasonable at the time of formation. Copyright assignment clauses are, as discussed, commonplace and hardly unconscionable terms. Because of the assignment, the new owner can use the data however and whenever they wish. It would almost certainly fail to argue that later computational use of workers' data retroactively renders a previous assignment unconscionable.

In the United States, unconscionability requires both procedural and substantive conditions to be met. Procedural unconscionability supposes that there has been oppression, that is inequality of bargaining power resulting in no real negotiation, and surprise, where the terms are hidden in prolix forms.[66] Substantive unconscionability asks whether the terms are one-sided and unduly harsh.[67] A contract that unreasonably favors the party asserting it is more likely to be found unconscionable as a matter of substance.[68] Compared to what is the position in the UK, the doctrine is notably broader in the United States and more favorable to workers. Still, American courts have generally been reluctant to strike down intellectual property assignment clauses on account of unconscionability, and for good reason. Such clauses are justified by the need to safeguard investment and flexibility in commercializing workers' work products, though there may be scope to challenge overly broad clauses where workers' bargaining power has been clearly exploited.[69]

The situation in continental Europe markedly diverges again. Unconscionability is not recognized as a doctrine on its own, and European contract law is subject only to minimum harmonization. The Unfair Terms in Consumer Contracts Directive 93/13/EEC sets out that a non-negotiated term is unfair if, contrary to good faith, it creates a significant imbalance in the parties' contractual rights and obligations to the consumer's detriment.[70] Yet this is of no assistance to workers, who fall outside the Directive's scope as they are not contracting as consumers. In many other respects, contract law is a matter of national law in Europe. Scandinavian contract statutes, for example, allow any contract to be voided if "unreasonable,"[71] while countries like Germany and France have narrower statutes, only providing protection against frustration, duress, usury, or contracts against public policy.[72]

A more recent development is the EU Data Act,[73] which harmonizes rules on data access and use between businesses, users and the public sector, with particular

66. *See* Zimmer v. CooperNeff Advisors, Inc., 523 F.3d 224, 228 (3d Cir. 2008); Nagrampa v. MailCoups, Inc., 469 F.3d 1257, 1280 (9th Cir. 2006).

67. *See* Pokorny v. Quixtar, Inc., 601 F.3d 987, 997 (9th Cir. 2010).

68. *Zimmer*, 523 F.3d at 228.

69. *See* Knuppel, *supra* note 44, at 985-95.

70. Council Directive 93/13, 1993 O.J. (L 95/29), at art. 3(1).

71. *See* CLAES-ROBERT VON POST, STUDIER KRING 36 § AVTALSLAGEN MED INRIKTNING PÅ RENT KOMMERSIELLA FÖRHÅLLANDEN 47 (1999).

72. *See* A.H. Angelo & E.P. Ellinger, *Unconscionable Contracts: A Comparative Study of the Approaches in England, France, Germany, and the United States*, 14 LOY. L.A. INT'L & COMP. L. REV. 455, 472–94 (1992).

73. Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on Harmonised Rules on Fair Access to and Use of Data and Amending Regulation (EU) 2017/2394

focus on facilitating data transfers within the Internet-of-Things framework.[74] It obliges "data holders" to provide data to "data recipients" under fair, reasonable, and non-discriminatory terms and in a transparent manner.[75] "Data" is defined broadly as any digital representation of acts, facts or information, including audiovisual material.[76] Article 13(1) renders unilaterally imposed terms concerning "access to and the use of data" by an enterprise on another enterprise not binding, if the terms are "unfair."[77] A term is unfair if its use grossly deviates from good commercial practice in data access and use, contrary to good faith and fair dealing.[78] Articles 13(4) and 13(5) set out a long yardstick list, although non-exhaustive, of circumstances where a contractual term may be considered unfair.[79] This includes where a party imposes terms allowing access to data in a way significantly detrimental to the other party's legitimate interests, particularly where trade secrets or intellectual property rights are involved.[80]

Although the EU Data Act was not drafted with employment or contractor relationships in mind, its broad definitions could potentially extend protection to employees and independent contractors. This hinges on whether they are considered an "enterprise" in a "business-to-business relation" with their employer or contractee. The term "enterprise" broadly covers any natural or legal person that is acting for purposes which are related to that person's trade, business, craft or profession.[81] Workers might be considered "enterprises" from that perspective, and if correct, then workers' data would also be considered "data" within the meaning of the Act.

Because of that, an overly broad copyright assignment clause, which is used to take full ownership of a worker's data to eventually replace their own labor, could be considered "unilaterally imposed" and govern the "access and use of the data [. . .] in a manner that is significantly detrimental to the legitimate interests of the other contracting party."[82] Again, Article 13(5)(b) specifically exemplifies that this may be the case where such data is protected by intellectual property rights.[83] Yet the application of the EU Data Act to workers' data is complicated by the fact that European copyright law is explicitly excluded from its scope.[84] However, and as discussed above, since European copyright law does not harmonize ownership and

and Directive (EU) 2020/1828, 2023 O.J. (L 2854) [hereinafter the "EU Data Act"] (on fair access to and use of data).

74. *Id.* at recital 5.
75. *Id.* at art. 8.
76. *Id.* at art. 2(1).
77. *Id.* at art. 13(1).
78. *Id*. at art. 13(3).
79. *Id.* at art. 13(4)–(5).
80. *Id.* at art. 13(5)(b).
81. *Id*. at art. 2(24).
82. *Id.* at art. 13(5)(b).
83. *Id.*
84. *Id.* at art. 1(8) (referring to Directives 2001/29/EC, 2004/48/EC, and the DSM Directive).

contractual assignments, these are left to Member States.[85] The question is therefore still left open, although the fact that European copyright law is excluded from the scope of the Act may suggest that, if the Act is given a purposive interpretation, copyright law specific to Member States should also be excluded.

C. WHETHER COMPANIES ARE UNJUSTLY ENRICHED BY THEIR WORKERS' DATA

Another potential remedy for workers is the doctrine of unjust enrichment, which is a common law cause of action. To succeed, the claimant must show: (i) that the defendant was enriched or received a benefit, (ii) at the claimant's expense, (iii) in circumstances that are unjust, and (iv) that there is no applicable defense.[86] Unlike tort, which compensates loss, unjust enrichment restores gains. It is, therefore, a claim of restitution and not compensation.[87]

Enrichment is assessed objectively at market value, though courts also recognize subjective devaluation.[88] Yet even if the parties may value a benefit differently, the claimant can overcome this if they prove that the defendant requested or freely accepted the benefit, or that the enrichment is an "incontrovertible benefit."[89] English courts have also held that for a benefit to be "at the claimant's expense," there does not need to be perfect equivalence between loss and gain, but rather a "sufficiently close causal connection."[90] Lastly, enrichment is considered "unjust" under English law when one of several recognized "unjust factors" is present, such as mistake, duress, undue influence, or a failure of basis.[91] Defenses, although controversial and not fully settled, include change of position, estoppel, and *bona fide* purchase for value, among others.[92]

Unjust enrichment as a body of law is recognized in other common law jurisdictions, including Australia, Canada, and the United States. The American approach similarly focuses on whether a defendant has received a benefit at the claimant's expense and whether it would be unjust to retain it without adequate compensation.[93] The Restatement (Third) of Restitution & Unjust Enrichment suggests the market value of the benefit as a key metric for assessing adequate

85. *See supra* Section III.

86. *See* Barton v. Morris [2023] UKSC 3 [77] (referring to Benedetti v. Sawiris [2013] UKSC 50, [¶ 10]) (UK).

87. *See* Banque Financière De La Cité v. Parc (Battersea) Ltd. [1998] UKHL 7 (UK).

88. *See* Benedetti v. Sawiris [2013] UKSC 50, [¶¶ 12–16] (UK).

89. *Id.* at ¶ 25.

90. GRAHAM VIRGO, THE PRINCIPLES OF THE LAW OF RESTITUTION 118 (3rd ed. 2015); *see also* Melissa Menelaou v. Bank of Cyprus UK Ltd. [2013] EWCA (Civ) 1960, [¶ 42] (UK); TFL Management Services Ltd. v. Lloyds Bank Plc [2013] EWCA (Civ) 1415, [¶ 82] (UK); Relfo Ltd. (in liquidation) v. Varsani [2014] EWCA (Civ) 360, [¶ 56] (UK).

91. *See* Deutsche Morgan Grenfell Grp. Plc v. Inland Revenue [2006] UKHL 49, [¶ 151] (UK).

92. Andrew Burrows, *Defences Arising from Changes in the Defendant's Circumstances*, in THE LAW OF RESTITUTION 521–34 (3rd ed. 2011).

93. *See* Colleen P. Murphy, *Recognizing Restitutionary Causes of Action and Remedies Under Rhode Island Law*, 20 ROGER WILLIAMS U. L. REV. 429, 432–4 (2015).

compensation, with other grounds, such as the claimant's cost, being used if no market exists.[94]

Applying this framework to a situation where an employer uses a worker's work products as data sources, it is clear that the employer is enriched by using the data for automation purposes. Meanwhile, the worker may gain little or even suffer loss if displaced as a result. However, the law of unjust enrichment is still not a perfect fit for the case of workers' data. Courts are generally reluctant to interfere with the terms of a freely negotiated contract, where the agreed-upon consideration has been paid. It is also a plausible defense that an employee's or contractor's compensation package represents fair market value for their services at the time of contracting. Any greater value extracted later would not necessarily be considered "unjust" for the lawful owner to keep.

These circumstances also do not fit squarely within the established "unjust factors." For the same reason that intellectual property assignments will generally not be unconscionable, nor will assignments with workers have been deemed entered into under undue influence.[95] There will also be a "basis" in the case of workers' data. This factor requires a total failure of the condition on which a benefit was conferred.[96] Workers enter into contracts for their services on the "basis" that they receive compensation for their work. This "basis" is received. In return, their employer or contractee receives the work product, and if copyright is assigned, then also acquires the ownership of any copyright forming part of such products. Workers do not enter contracts on the basis that they will be further compensated for their employer's later use of their work products, such as for automation. The fact that copyright has been intentionally assigned reinforces this assumption. While different common law jurisdictions may analyze these issues somewhat differently, it suffices to say that workers will face considerable difficulty in relying on unjust enrichment for restitution.

D. EMPLOYMENT LAW RESTRICTIONS OF REDUNDANCIES AND USING WORKERS' DATA

It is possible that, in some cases, using workers' data to automate workstreams will result in those same workers becoming redundant, in whole or in part, for some of their tasks. It falls outside the scope of this Article to discuss in more detail whether companies may, as a matter of employment law, terminate employees on grounds of redundancy in these circumstances. Broadly speaking, employees will have little to no say on the matter in most states in the United States, which allow employers to hire and dismiss their employees "at will" without having to establish

94. RESTATEMENT (THIRD) OF RESTITUTION & UNJUST ENRICHMENT § 49(3) (A.L.I. 2011).

95. *See supra* Section IV.B.

96. Fibrosa Spolka Akcyjna v. Farburn Lawson Combe Barbour Ltd. [1943] AC 32, [¶ 77] (HL) (UK); Rover Int'l Ltd. v. Cannon Film Sales Ltd. [1989] 1 WLR 912 ,[¶ 924b] (UK); Stocznia Gdanska SA v. Latvian Shipping Co. [1998] 1 WLR 574, [¶ 588] (UK).

just cause.[97] Employment laws provide greater protection for employees in the UK and continental Europe. Although not harmonized across Europe, employment laws typically require employers to justify redundancies, follow proper procedural requirements, and use objective selection criteria.[98] Taking the UK as an example, the Employment Rights Act 1996 defines two situations where a genuine redundancy may occur, including where (i) the employer ceases or intends to cease to carry on the business for the purposes of which the employee was employed, and (ii) the business requirement for the employee to carry out work of a particular kind ceases or diminishes.[99]

There is a long history of employers optimizing workstreams and patterns to create efficiencies within their operations and workforces, which sometimes lead to redundancies. It is well-recognized that there is a genuine redundancy where the employer reorganizes the business to improve its efficiency, meaning that fewer people are needed to do the same amount of work. This is a question of fact, not law.[100] If there is no longer a need for the work to be done, the work is redundant. A job may also become redundant where a new process or system is introduced, if this causes the work to no longer be needed. In the UK, in deciding the question whether a redundancy situation exists, courts are not concerned with deciding whether the employer acted with commercial good sense.[101] It is not a requirement that the dismissal of the employee actually achieves a cost saving for the employer.[102] English courts will also not go behind the facts and investigate how the redundancy situation arose and whether it could have been avoided and whether there are any viable alternatives,[103] as the cessation or diminution of the work previously required may arise "for whatever reason."[104] If an employer has optimized its business operations under similar circumstances, for example by adopting new AI technologies, it is possible that there could be a genuine redundancy situation. This position is unlikely to change even if such AI technologies have been developed with employees' data. Employers may however be required to make a proper assessment of whether the employee, made redundant by AI, can

97. *See* Katherine V.W. Stone, *Dismissal Law in the United States: The Past and Present of At-Will Employment*, INT'L. COLLABORATIVE ON SOC. EUR. (2007).

98. *See, e.g.,* Rochelle Le Roux & Allison Fiorentino, *Redundancy and Judicial Power: Between Inaction and Creative Boldness*, 4 COMP. LAB. L. & SOC. SEC. J. 126 (2017) (summarizing redundancy laws across Europe).

99. Employment Rights Act 1996, c. 18, § 139(1) (UK).

100. *See* Murray v. Foyle Meats Ltd. [1999] ICR 827, [¶ 829] (HL) (appeal taken from N. Ir.) (UK) (concluding that the diminution in the employer's need for its employees was a question of fact, not law).

101. *See* Moon v. Homeworthy Furniture (Northern) Ltd. [1977] ICR 117 (EAT) (UK); James W Cook (Wivenhoe) Ltd. v. Tipper [1990] ICR 716 (EAT) (UK).

102. *See* Jones v. BT Facility Servs. Ltd. [2020] EAT (appeal no. 0237/19/BA) (UK).

103. *See* TNS UK Ltd. v. Swainston [2014] EAT (appeal no. 0603/12/BA) (UK) (citing H. Goodwin Ltd. v Fitzmaurice [1977] IRLR 393 (EAT) (UK); Moon v. Homeworthy Furniture (Northern) Ltd. [1976] IRLR 298, [1977] ICR 117 (EAT) (UK); Ass'n of Univ. Tchrs. v. Univ. of Newcastle-upon-Tyne [1987] ICR 317 (EAT) (UK); James W Cook & Co (Wivenhoe) Ltd. v. Tipper [1990] IRLR 386, [1990] ICR 716 (CA)(UK)).

104. Employment Rights Act 1996, c. 18, § 136(6) (UK).

be redeployed in other areas and the failure to do so could still potentially result in unfair dismissal claims.[105]

E. WHAT AI REGULATIONS SAY ABOUT USING WORKERS' DATA

In May 2024, the EU adopted its new EU AI Act, which sets out comprehensive obligations for the development, supply and deployment of AI systems and models.[106] Providers of high-risk AI systems face the most stringent requirements, including establishing risk management systems, drawing up technical documentation throughout the system's life cycle, putting in place record-keeping and transparency procedures, and ensuring accuracy, cybersecurity, and quality management.[107] Article 10 sets specific rules for data governance, including what testing datasets may be used for high-risk AI systems, and that providers must take measures to mitigate biases in datasets.[108] In addition, high-risk AI systems must undergo a conformity assessment procedure through internal control or responsible notified bodies,[109] register in public databases,[110] and conform to certain reporting obligations.[111] Deployers of high-risk AI systems have fewer obligations to follow, but must, for example, assign human oversight to personnel, monitor the operation of the high-risk AI system on the basis of the instructions for use, and keep automatically generated logs.[112] Providers of general-purpose AI models must put in place a copyright compliance policy,[113] provide certain technical documentation,[114] and provide summaries of training data.[115] Providers of general-purpose AI models with systemic risk[116] are subject to more onerous requirements.[117] Companies that provide or deploy AI systems that are neither

105. Williams v. Compair Maxam Ltd. [1982] ICR 156 (EAT) (UK) (establishing the five redundancy criteria under English employment law, including the fifth criterion that employer shall seek to see, if instead of making employees redundant, they could offer alternative employment to them instead).

106. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024, Laying Down Harmonised Rules on Artificial Intelligence, 2024 O.J. (L 1689) [hereinafter, the "EU AI Act"].

107. *Id.* at arts. 9, 11–13, 15, 17–18.

108. *Id.* at art. 10.

109. *Id.* at arts. 16(f), 43.

110. *Id.* at art. 49.

111. *Id.* at art. 20.

112. *Id.* at art. 26.

113. *Id.* at art. 53(1)(c).

114. *Id.* at art. 53(1)(a)–(b).

115. *Id.* at art. 53(1)(d).

116. *See id.* at art. 3(65) (defining "systemic risk" as: "A risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain.").

117. *Id.* at art. 51(1).

high-risk nor general-purpose need to comply with fewer requirements, such as transparency,[118] and staff AI literacy.[119]

It is expected that other countries will follow the footsteps of the EU AI Act, adopting their own AI regulations, and in particular countries which did so for data privacy after the GDPR was enacted. For example, Brazil[120] and Canada[121] have announced similar draft AI regulations that introduce requirements for AI systems based on certain risk thresholds. The United States has no federal AI regulations as of yet, and with no federal initiatives pending, individual states are taking the lead. Colorado became the first state in May 2024 to adopt specific AI regulations, which take a staggered risk-based approach similar to the EU AI Act. [122] Texas also recently passed its own set of AI regulations, the Texas Responsible Artificial Intelligence Governance Act.[123]

The EU AI Act excludes employment law from the scope of the Act, expressly leaving Member States free to introduce provisions more favorable to workers.[124] This does not mean, however, that workers' rights are not impacted, at least indirectly. Deployers of high-risk AI systems must conduct a fundamental rights impact assessment prior to deployment.[125] This requires, *inter alia*, identifying (c) the categories of natural persons or groups likely to be affected, (d) the risks of harm they face, and (f) the measures to be taken should such risks materialize, including governance arrangements and complaint mechanisms.[126] Deployers must then notify the market surveillance authority of the results.[127] Where high-risk systems threaten to replace workers, deployers would be obliged to identify those workers as affected persons and specify the harm. Redundancy arguably constitutes such "harm." However, and importantly, Article 27 merely requires that deployers assess what measures they *should* take, such as internal governance and complaint mechanisms, without requiring that they adopt any particular measure.[128] There would be no obligation for deployers to issue compensation to workers made redundant due to AI. These protections are further limited in scope in that most AI systems adopted to automate workstreams are unlikely to fall within the high-risk category.

118. *Id.* at art. 50.
119. *Id.* at art. 4.
120. Senado Federal, Bill No. 2338/2023 (Dispõe sobre o uso da Inteligência Artificial [Framework for Artificial Intelligence]), Sessão Legislativa de 2023 (Braz.).
121. Artificial Intelligence and Data Act, Bill C-27, pt. 3, 44th Parl., 1st Sess., (Can. 2022).
122. S.B. 24-205, 2024 74th Gen. Assemb., Reg. Sess. (Colo. 2024).
123. H.B. 1709, 2024 89th Gen. Assemb., Reg. Sess. (Tex. 2024).
124. EU AI Act, *supra* note 106 at art. 2(11).
125. *Id.* at art. 27(1).
126. *Id.*
127. *Id.* at art. 27(3).
128. *Id.* at art. 27.

V. THE CASE FOR COMPANIES' RIGHT TO USE WORKERS' DATA

Adopting new technologies to improve efficiency, even at the cost of redundancies, is nothing new. Companies have a legitimate interest in leveraging innovation to remain competitive, enhance productivity, and increase shareholder value. This process of "creative destruction" is a core feature of capitalism, where outdated practices give way to more efficient alternatives.[129] Historically, technological change has reduced workforces, justified by companies as necessary for their long-term success. AI is no different.

Company directors are under a statutory duty in many jurisdictions to act in good faith to promote the success of the company for the benefit of its members, typically understood as the shareholders.[130] In the UK, this is encapsulated in the principle of enlightened shareholder value.[131] Success is generally measured by profitability, growth, and returns, but not equated with maximizing profits in the abstract. Directors enjoy wide discretion in this regard, provided their judgment is exercised in good faith.[132] In practice, this may include adopting AI technologies that displace human labor where they can perform tasks more accurately, quickly or cost-effectively. Corporate law imposes no parallel duty to protect employees or contractors.

Because companies often are the lawful owners of workers' data under copyright laws, or contractual arrangements, the premise is that they are free to use that data. This remains the case even if workers are harmed or made redundant by its use, and even though such data would not exist without the worker's contribution in the first place.[133] From the company's perspective, this unavoidable account is justified. It invests in employees by providing training, tools, and remuneration, and in return acquires the outputs produced as consideration for their own bargain. Companies have a legitimate interest in obtaining copyright ownership over their workers' work products. Assignment gives the company exclusive control to use, modify, commercialize, and integrate works into larger projects, without concerns about joint ownership or competing claims. It also ensures continuity of ownership when workers depart and prevents former employees from licensing works to competitors. While licensing could achieve much of this, assignment provides greater clarity and security.

The real complaint seems to be not that workers' data is used by companies, but that workers are not necessarily adequately compensated for use, and particularly

129. Faruk Ulgen, *Creative Destruction*, *in* ENCYCLOPEDIA OF CREATIVITY, INVENTION, INNOVATION AND ENTREPRENEURSHIP 1–8 (Elias G. Carayannis ed., 2017).

130. *See, e.g.,* Companies Act 2006, c. 46, §172 (UK). In the United States, corporate law is governed primarily by state law, with Delaware being most influential. Under Delaware law, directors owe fiduciary duties to the corporation, including the duty of care and the duty of loyalty. The duty of loyalty requires directors to act in good faith to advance the best interests of the corporation and its stakeholders. *See* In re Oracle Corp. Derivative Litig., 824 A.2d 917, 920 (Del. Ch. 2003).

131. Re Smith and Fawcett Ltd [1942] Ch 304, [¶ 306] (UK).

132. *Id.*

133. *See supra* Section III.

where those outputs later underpin AI systems deployed against their own interests. Yet contract law requires only that consideration be sufficient, not equal or adequate.[134] From a company's perspective, the status quo is a natural extension of ownership and contractual freedom: workers have been paid for their labor, and any data generated in the course of employment becomes company property, which workers have no say about.

VI. THE CASE FOR WORKERS' RIGHT TO REDRESS

The current framework of contract and employment law, which prioritizes freedom of contract and corporate ownership of workers' data, is not without criticism. In theory, it allows some companies to hire workers merely to capture their data, automate the same tasks, and then make those workers redundant. Unless otherwise agreed, and as discussed, workers are entitled to no additional compensation under contract or employment law when their data is used against their own interests.[135]

Data is one of the most valuable resources in the age of AI, including for corporate automation purposes.[136] Workers provide not only tangible work outputs but also high-quality data sources. While they are paid for their immediate contributions, the value of their work increases when repurposed as data for automation. That added value, realized through productivity gains and cost savings, currently flows to company shareholders and executives, which calls into question the balance of power in employer-employee and contractee-contractor relationships. Ultimately, it invites the broader question of whether benefits from AI development should be more equitably shared with those who contribute to the underlying work products.

The dilemma is similar to that of the debate surrounding the development of AI models, particularly foundation models and large language models ("LLMs"), some of which were advanced using vast amounts of publicly available (often copyrighted) materials, and allegedly sometimes without permission from rightsholders.[137] Although LLMs do not replace authors or their works directly, they may reduce demand for them.[138] Yet these rightsholders are in a stronger legal position than workers. Unlike workers, they retain copyright and can claim compensation for infringement. In the EU, where text and data mining is subject to an opt-out regime, they may also preclude their works from being used in training.[139] By contrast, workers who assign copyright to their employers, or whose

134. *See* Chappell & Co Ltd v. Nestle Co Ltd [1959] UKHL 1.

135. *See supra* Section IV.

136. *See* Wang, Prabhat & Sambasivan, *supra* note 8.

137.*See, e.g.*, Kadrey v. Meta Platforms, Inc., No. 23-CV-03417-VC, 2025 WL 1752484, at *1 (N.D. Cal. June 25, 2025).

138.*See* Mattias Rättzén, *Location Is All You Need: Copyright Extraterritoriality and Where to Train Your AI*, 26 COLUM. SCI. & TECH. L. REV. 175, 200–01 (2024) (discussing, in the context of U.S. fair use arguments, how AI could reduce the demand for authors' original works).

139. *See* DSM Directive, *supra* note 10 at arts. 4(1), 4(3).

rights vest automatically by statute, are no longer "rightsholders" and cannot opt out.

Workers who "sweat the brow" for their employers are therefore worse off than external rightsholders. Once copyright is assigned, workers lose control over future uses of their works, and assignees owe no duty of disclosure. Secondary uses, such as data aggregation or machine learning, could in certain cases generate greater economic value than the original outputs, yet none of this value returns to workers.[140] There is an inherent lack of transparency. Workers may, in a misinformed state, assume that their companies are only using their work products tangibly, only to later find out that they have been accumulated as data fuel for automation purposes. Some might refuse to work for such companies, while others may demand additional compensation if they know their outputs serve broader purposes that could ultimately displace them in the future.

Freedom of contract is of little help to workers here. As discussed above, copyright is assigned to employers by default in several jurisdictions like the UK and the United States.[141] Although workers could in theory resist assignment or seek to limit secondary use, they usually lack bargaining power.[142] Copyright assignments for work products are routinely boilerplate clauses, and most companies would refuse to engage workers without owning the rights. While there are sound reasons for this, the imbalance suggests that if workers' data is to be treated more equitably, it is plain that statutory intervention is necessary.[143]

VII. THE EUROPEAN "BEST-SELLER" RULE AND FAIR REMUNERATION FOR COPYRIGHT ASSIGNMENTS

A. THE "REAL" VALUE GAP IN THE DSM DIRECTIVE

One of the most controversial reforms in EU copyright law in recent years was the adoption of the DSM Directive.[144] The DSM Directive introduced a right for authors and performers to receive appropriate and proportionate remuneration for the exploitation of their works, along with the ability to adjust existing contracts to achieve this.[145] The reform addressed concerns that authors and performers, typically in a weaker bargaining position, may lack information on how their works are exploited and what revenues they generate.[146] This is problematic as it prevents them from assessing the true value of their rights, or from comparing different

140. *See infra* Section VIII (discussing in more detail the prevalence of a so-called "workers' data value gap").

141. *Supra* Section III.

142. *See* WOUTER ZWYSEN, MONOPSONY AND NON-COMPETITIVE LABOUR MARKETS: WORKERS' WEAKENING BARGAINING POSITION, EUR. TRADE UNION INST. (Nov. 2024).

143. *See supra* Section V.

144. DSM Directive, *supra* note 10.

145. *Id.* at art. 20.

146. *Id.* at recital 72; *see also Commission Impact Assessment on the Modernisation of EU Copyright Rules*, at 173–75, SWD (2016) 301 final.

licensing deals.[147] As the European Commission observed in its impact assessment, creators in the music and audiovisual sectors particularly struggle with online and cross-border exploitation, where remuneration often bears little relation to use or success.[148] The rationale is that if remuneration to rightsholders is meant to reflect the economic value of their rights, then a framework should be established to facilitate the exchange of information, especially where they otherwise lack the necessary insight to assess that value, enabling them to negotiate fair remuneration. It was against this backdrop that what are now Articles 18–20 of the DSM Directive, also described as the "real" value gap, were negotiated and eventually introduced.

Article 18(1) entitles authors and performers to receive "appropriate and proportionate remuneration" when exploiting their works.[149] Recital 73 clarifies that this should reflect the actual or potential economic value of the licensed rights, taking into account the creator's contribution to the overall work, and all other circumstances of the case, such as market practices or the actual exploitation of the work.[150] A lump sum payment can also constitute proportionate remuneration, but should be the exception rather than the norm.[151] When implementing Article 18(1), Member States are free to use different legal mechanisms to take into account the principle of contractual freedom and a fair balance of rights and interests.[152] Sweden, for example, has gone one step further, such that any contractual terms which restrict the right to fair remuneration are automatically void.[153]

Lack of transparency lies at the root of the problem of what constitutes fair remuneration. If rightsholders cannot ascertain how their works are being used, then they cannot ascertain whether or not they are fairly compensated. With that in mind, Article 19 requires transferees, or their successors, to provide regular, relevant, and comprehensive information on exploitation, revenues, and remuneration.[154] This transparency obligation must be effective yet proportionate in every sector, allowing Member States to limit it where administrative burdens would outweigh the revenues at stake or where the creator's contribution is insignificant, unless information is necessary for contract adjustment.[155]

Still, a statutory right to fair remuneration does not automatically alter existing contracts. Because creators are typically in a weaker bargaining position, their ability to renegotiate contracts can be limited. Authors and performers are also often reluctant to enforce their rights against their contractual partners before a

147. DSM Directive, *supra* note 10 at recital 72; *Commission Impact Assessment*, *supra* note 146, at 174–75.

148. *Commission Impact Assessment*, *supra* note 146, at Annex 2, 30.

149. DSM Directive, *supra* note 10, at art. 18(1).

150. *Id.* at recital 73.

151. *Id.*

152. *Id.* at art. 18(2).

153. Prop. 2021/22:278, *supra* note 40 at 169–70.

154. DSM Directive, *supra* note 10, at art. 19.

155. *Id.* at art. 19(3)–19(4).

court or tribunal.[156] It was against that backdrop that Article 20 introduced a statutory contract adjustment mechanism, also known as the "best-seller" clause.[157] The "best-seller" clause reads:

> "Member States shall ensure that, in the absence of an applicable collective bargaining agreement providing for a mechanism comparable to that set out in this Article, authors and performers or their representatives are entitled to claim additional, appropriate and fair remuneration from the party with whom they entered into a contract for the exploitation of their rights, or from the successors in title of such party, when the remuneration originally agreed turns out to be disproportionately low compared to all the subsequent relevant revenues derived from the exploitation of the works or performances."[158]

Article 21 supplements this by introducing a voluntary, alternative dispute resolution mechanism for any contract adjustment or transparency claims: "Member States shall also ensure that representative organizations of authors and performers may initiate such procedures at the specific request of one or more authors or performers."[159]

The Directive's initial draft contained only a contract adjustment mechanism where the agreed remuneration turned out to be "disproportionately low".[160] This was eventually split up into two, what is now Articles 18 and 20, following suggested amendments from the JURI Committee, to ensure that creators also receive fair remuneration at the outset of new contracts.[161] The rationale was to empower creators who are vulnerable in negotiations by guaranteeing equitable remuneration and enabling adjustment where returns otherwise prove unfair.[162] The logic, therefore, is that creators shall be entitled to fair remuneration when negotiating new contracts for the exploitation of their works. Creators shall also be provided with the necessary information to subsequently evaluate those contracts and, if their agreed remuneration turns out to be disproportionately low, be entitled to adjust them accordingly.

The right to claim additional remuneration is particularly significant in long-term exploitation contracts where rights increase in value after conclusion.[163] However, the wording of the article is not limited to any such case. Indeed, it was pointed out on multiple occasions during the legislative proceedings that the drafting language was very broad.[164] Articles 18 and 20 apply to all "authors" and "performers" for

156. *See id.* at recital 79.

157. *Id.* at art. 20; *see also Commission Impact Assessment*, *supra* note 146, at 187.

158. DSM Directive, *supra* note 10, at art. 20(1).

159. *Id.* at art. 21.

160. *Proposal for a Directive on Copyright in the Digital Single Market*, art. 15, COM (2016) 593 final [hereinafter the "DSM Proposal"].

161. *Draft Report on the Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market*, PE 601.094v01-00, at 43–44 (Mar. 10, 2017).

162. *Id.*

163. *See* DSM Directive, *supra* note 10, at recital 78.

164. *See* LIONEL BENTLY ET AL. (EUR. PARL. COMM. LEGAL AFFAIRS), STRENGTHENING THE POSITION OF PRESS PUBLISHERS AND AUTHORS AND PERFORMERS IN THE DSM DIRECTIVE 48 (Sept. 2017).

the "exploitation" of their works or rights, which are terms left undefined.[165] Although Article 18 was negotiated with online exploitation in mind, its neutral language means that it is expected to be equally neutrally implemented into national law. Importantly, Article 20 also contains no time limit, which exposes exploiters to potential retroactive claims down the line.[166] That risk, however, can be mitigated by providing transparent information under Article 19 and ensuring fair remuneration at the initial contracting stage.

Contract adjustment mechanisms in copyright law are nothing new in Europe. Several Member States, including Germany, Belgium, France, the Netherlands, and Spain, already had such provisions before the DSM Directive.[167] Indeed, several of those national rules served a basis for the discussions when the DSM Directive was negotiated, with Germany being most notable, having had such a mechanism since the adoption of the German Copyright Act 1965.[168] The Act provided that where an author's remuneration was "conspicuously disproportionate" to the proceeds and benefits derived from use, the author could demand modification of the agreement to secure equitable participation.[169] The right applied even if remuneration was fair at the time of contracting, regardless of foreseeability, and extended to all works.[170]

Such mechanisms historically had limited market impact.[171] Rightsholders often hesitated to enforce them for fear of retaliation, contract termination, or market exclusion or "blacklisting," while litigation costs likely had a further deterrent effect.[172] However, there is lack of data from Member States on industry uptake, making it difficult to draw any real conclusions on how effective the rules are. What we know is that Germany has experienced many of the cases. A notable such example is the German Federal Court of Justice's ruling from 2011 involving the cameraman of the film *Das Boot*.[173] The Court held that a "conspicuous

165. DSM Directive, *supra* note 10, at arts. 18, 20.

166. *Id.*

167. For a comparative survey, see *Commission Impact Assessment*, *supra* note 146, at Annex 14D.

168. UrhG [Gesetz über Urheberrecht und verwandte Schutzrechte] [Copyright Act], Sept. 9, 1965, BGBL. I at 1273 (Ger.) [hereinafter "German Copyright Act 1965"].

169. *Id.* at § 32a.

170. *Id.*; *see also* BENTLY ET AL., *supra* note 164, at 62–63.

171. Bently et al. describe it as that "[t]he economic evidence is that bestseller clauses are of limited effect and often merely assist those who have already acquired the market power to renegotiate existing deals." *Id.* at 84. Other scholars have similarly argued that the mechanism is "inapt at solving the more general issue of uneven bargaining power." *See* DUSOLLIER ET AL., *supra* note 38, at 71. Dutch scholars have also noted that the measure in the Netherlands has been "rarely being invoked in practice." *See* Stef J. van Gompel et al., *Evaluatie Wet Auteurscontractenrecht: Summary and Conclusions*, UNIV. LEIDEN 1, 4 (Sep. 1, 2020), https://www.ivir.nl/publicaties/download/Evaluatie-ACR-Eindrapport-Summary-and-conclusions.pdf [https://web.archive.org/web/20250926184240/https://www.ivir.nl/publicaties/download/Evaluatie-ACR-Eindrapport-Summary-and-conclusions.pdf].

172. van Gompel et al., *supra* note 171, at 4.

173. Bundesgerichtshof [BGH] [Federal Court of Justice], I ZR 127/10, Das Boot (Sept. 22, 2011) (Ger.); *see also* Amélie Lacourt, Justine Radel-Cormann & Sophie Valais, *Fair Remuneration for Audiovisual Authors and Performers in Licensing Agreements*, EUR. AUDIOVISUAL OBSERVATORY: IRIS PLUS 1, 87 (Dec. 2023).

disproportion" exists where the agreed remuneration is less than half of what is deemed equitable, which brought the cameraman further payment of €462,000, on top of the €104,303 lump sum he had originally been paid for his work.[174] The term "conspicuous disproportion" in the German Copyright Act 1965 has since been replaced with "disproportionately low remuneration" to reflect the wording of the DSM Directive.[175] The statutory amendment retriggered litigation between the parties in *Das Boot*, but was eventually settled.[176]

A related, but different mechanism in Member States' copyright laws are measures that address new uses of copyrighted works. Parties may have initially agreed to a lump sum or royalties based on the assumption that a work would be used in a specific way. However, changes in market circumstances or new technology could enable different uses, rendering the initial remuneration unreasonable. Once again, Germany has led regulations in this area. For many years, the German Copyright Act 1965 forbade licensing rights for "unknown means of utilisation."[177] This meant that for any new use, an author could choose to extend their license and renegotiate payment.[178] But as new technologies emerged more frequently, this rule became ineffective. It was replaced in 2008 by a less rigid measure requiring authors to grant licenses or assignments for "unknown types of use."[179] If a licensee or assignee wants to exploit a work for a new, unstated use, they must notify the author, who is entitled to refuse the grant.[180] Similar "new use" provisions have since been adopted in the Netherlands and France.[181] The Dutch version specifically allows the author to claim "additional fair compensation" for uses that were unknown when the contract was made.[182] Since then, in both Germany and the Netherlands, these provisions have been interpreted narrowly to only apply to innovations that are transformative from the point of view of

174. Karl-Nikolaus Peifer, *Die unendliche Geschichte der Nachvergütung von Miturhebern im Filmbereich [The Never-ending Story of Additional Remuneration for Co-authors in the Film Sector]*, ZUM 814 (2021) (Ger.).

175. UrhG [Gesetz über Urheberrecht und verwandte Schutzrechte], as amended by Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarkts, May 31, 2021, BGBL. I at 1204 (Ger.) [hereinafter "German Copyright Act 2021"], § 32a (in English translation: "[w]here the author has granted to another a right of use on conditions which, taking into account the author's entire relationship with the other party, result in the agreed remuneration proving to be disproportionately low in comparison to the proceeds and benefits derived from the use of the work, the other party is obliged, at the author's request, to consent to a modification of the agreement which grants the author further equitable participation appropriate to the circumstances. It is irrelevant whether the parties to the agreement had foreseen or could have foreseen the amount of the proceeds or benefits obtained.").

176. Lacourt, Radel-Cormann & Valais, *supra* note 173, at 88.

177. German Copyright Act 1965, *supra* note 168, at § 31(4).

178. U.K. INTELLECTUAL PROP. OFF. (U.K. IPO), RIGHTS REVERSION AND CONTRACT ADJUSTMENT (Feb. 6, 2023), at ch. 3.2.2.

179. German Copyright Act 2021, *supra* note 175, at § 31(a).

180. *Id.*

181. Lucie Guibault & Bernt Hugenholtz, *Study on the Conditions Applicable to Contracts Relating to Intellectual Property in the European Union*, INST. FOR INFO. L., No. ETD/2000/B5-3001/E/69, at 51–52, 67 (May 2002).

182. Dutch Copyright Act, *supra* note 41, at § 25c(6).

consumers and which have the capability to open up new markets.[183] Although a "new use" provision was recommended in preparatory European policy reports, it was not included in the final draft of the DSM Directive.[184] This is understandable, given the broad scope of the right to fair remuneration and best-seller rule in the DSM Directive, which address any inequitable economic distribution, regardless of the cause.

Article 26(2) of the DSM Directive provides that it applies without prejudice to acts concluded and rights acquired before its entry into force.[185] Sweden has interpreted this to exclude earlier contracts, but Member States are free to deviate from that when implementing the rules into national law.[186] Sweden did precisely that, and extended the "best-seller" rule retroactively to contracts concluded within the previous twenty years, reasoning that the mechanism is designed to protect creators from inequitable bargains.[187] Some Member States, like the Netherlands, have already and previously applied their contract adjustment mechanisms retroactively.[188] Other Member States, like Germany, only extend their legislation to contracts that were entered into before its entering into force.[189]

B. FAIR REMUNERATION FOR WORKERS' DATA

In the draft proposal for the DSM Directive, the right to fair remuneration and "best-seller" rule for contract adjustments were neutrally drafted for all types of exploitation contracts.[190] Sweden, which has a long history of collective agreements on the labor market and strong unions, proposed that the rules should expressly exclude employment contracts. Other Member States did not share Sweden's standpoint.[191] Instead, they pointed out that excluding certain types of contracts from the scope of the rules could make it easier for parties to avoid the rules.[192] The draft text in the Directive did not change in scope. In light of that history,

183. *See* U.K. IPO, *supra* note 178, at ch. 3.2.2 (referring to Artur-Axel Wandtke & Henrik Holzapfel, *Ist § 31 IV UrhG noch zeitgemäß*, GRUR 2004, [¶ 290]); *see also* Guibault & Hugenholtz, *supra* note 181, at 79 (stating, with respect to the German Copyright Act 1965, that "'future or unknown forms of exploitation' have been defined as those forms of use that are not technically possible or, even if so, the economic relevance of which is not known at the time of conclusion of the contract. Hence, a form of use is deemed to be new, when it is a clearly distinguishable economic and technical mode of exploitation of a work.").

184. UK IPO, *supra* note 178, at ch. 3.2.2 (referring to DUSOLLIER ET AL., *supra* note 38, at 104).

185. DSM Directive, *supra* note 10, at art. 28(2).

186. Prop. 2021/22:278, *supra* note 40, at 162–64.

187. *Id.* at 180.

188. *See* U.K. IPO, *supra* note 178, at ch. 3.2.1.

189. *Id.*

190. *See* DSM Proposal, *supra* note 160, at art. 15 (worded as, ". . .are entitled to request additional, appropriate remuneration from the party with whom they entered into a contract for the exploitation of the rights when the remuneration originally agreed is disproportionately low compared to the subsequent relevant revenues and benefits derived from the exploitation of the works or performances").

191. Prop. 2021/22:278, *supra* note 40, at 159.

192. *Id.*

Sweden concluded, when implementing the Directive into national law, that it was clear that the rules apply to all assignments, including those in employment contexts.[193] This was considered to apply regardless of whether the copyright assignment was express or implied, and whether the employment contract and its terms were agreed in writing or orally.[194] However, Sweden considered that neither Article 18 nor Article 20 are intended to apply to contracts entered into with organizations or unions, which are not directly associated with any contract with an individual author or performer.[195] If that assumption is correct, then some collective agreements on the labor market are likely to be excluded from the scope of the rules.[196]

Although the rules on fair remuneration and the "best-seller" rule were originally intended for a different context, they would also apply to employment and contractor relationships where workers create copyright-protected content. The "best-seller" rule in Article 20 extends not only to exploitation licenses but also to copyright assignments, which are commonly found in employer-employee and contractor agreements.[197] In practice, this means the rule could benefit individuals whose creative works generate significant revenue for the organizations they serve, yet who are not compensated proportionally. Yet there is another important aspect which has to be considered as well, which is the data accumulated from such work products. Works produced by workers are increasingly valuable for automation purposes, as previously discussed.[198] When entering into employment or independent contractor agreements, where copyright may be assigned, workers are frequently not told how their work products will end up being used. The economic value of those work products could be increased when they are additionally used as data, for example for training AI systems, and especially if those systems are later deployed to replace their jobs.[199] Many workers have so far not received any additional compensation to account for that use and extracted added value. Whether the compensation agreed upon at the outset of the agreement is fair and proportionate will depend on the specific circumstances.

193. *Id.* at 159–60.

194. *Id.* at 160.

195. *Id.*

196. The DSM's recital 72 appears to confirm this, stating "[t]hat need for protection does not arise where the contractual counterpart acts as an end user and does not exploit the work or performance itself, which could, for instance, be the case in some employment contracts." DSM Directive, *supra* note 10, at recital 72.

197. *See id* at recital 78 ("Accordingly, without prejudice to the law applicable to contracts in Member States, a remuneration adjustment mechanism should be provided for as regards cases where the remuneration originally agreed under a *license or a transfer of rights* clearly becomes disproportionately low compared to the relevant revenues derived from the subsequent exploitation of the work or fixation of the performance by the contractual counterpart of the author or performer." (emphasis added)).

198. *See supra* Sections II and V.

199. *See infra* Section VIII (discussing in more detail the existence of a so-called workers' data value gap).

Extending the "best-seller" rule to workers' data when used for developing and deploying AI solutions would more equitably reward workers for their work, especially where such use may ultimately contribute to job displacement. But assigning a price on that data is far from straightforward. The fundamental question is: at what point does the value of workers' outputs, when used as data, increase so substantially that the originally agreed compensation becomes "disproportionately low"? This gives rise to several related questions, including whether the economic value of a worker's output should be assessed solely in relation to the individual work itself, or whether it should also reflect the work's subsequent use, particularly its value when repurposed as data. Article 18 and Recital 73 of the DSM Directive suggest the latter. Article 18(1) ensures that authors receive fair remuneration for the exploitation of their works "or other subject matter," and Recital 73 explains that such economic value should be both "actual or potential" and take into account "the author's or performer's contribution to the overall work or other subject matter and all other circumstances of the case, such as market practices or the actual exploitation of the work."[200] This indicates that assessing economic value can include the use of copyrighted works as data.

This raises further challenges about how to value creative output when its worth is amplified not individually, but in aggregate. For instance, what is the economic value of a single work product that has limited standalone significance, but becomes substantially more valuable when pooled with others? Data typically has little value in isolation but becomes exponentially more valuable when aggregated. A purposive reading of Recital 73 would support evaluating economic value in light of this broader context. Consider a hypothetical: if ten documents are worth €100, but fifty similar documents used as training data are worth €1,000, fairness would suggest that remuneration should be calculated based on the latter value, if such large-scale exploitation occurs. Similarly, assume that a graphic designer who creates 500 images for an employer receives €10,000 for the set. Yet if those images are used to train an AI system capable of producing infinite variations of the same style, the economic value of the original images may be far greater than initially agreed. Articles 18 and 20 are designed to account for such scenarios by allowing for reassessment when new factors influence the fairness of remuneration.[201] There is no indication that the use of copyrighted works as data in AI development should be excluded from that analysis.

The extension of the DSM Directive to workers' data would have far-reaching consequences. Although the Directive requires implementation through national legislation among Member States, Article 18 offers workers, whether employees or independent contractors, a legal tool to negotiate fair compensation for the data their work generates.[202] Under Article 19, employers and contractees are obliged to provide "relevant and comprehensive information" about the exploitation of

200. DSM Directive, *supra* note 10, at recital 73, art. 18.
201. *Id.* at arts. 18, 20.
202. *Id.* at art. 18.

works on a regular basis.[203] In practice, this requires disclosing whether and how a particular worker's data is being used for automation or AI-related purposes. If a worker has received compensation for their work products but not for their subsequent use as data, they may be entitled under Article 20 to claim additional, fair remuneration. This right is triggered when "the remuneration originally agreed turns out to be disproportionately low compared to all the subsequent relevant revenues derived from the exploitation of the works or performances."[204] As the use of workers' data to train AI systems becomes increasingly common, the importance of transparency, accountability, and equitable remuneration will only grow. The DSM Directive provides a flexible framework that could help ensure that workers are not left behind in this data-driven economy, although each case will ultimately depend on the specific facts.

C. A CONTRACT ADJUSTMENT ONLY, NOT LAW ADJUSTMENT

A potentially significant limitation of the "best-seller" rule under the DSM Directive is that it only permits the adjustment of contracts, not of the legal default ownership of rights. This distinction is essential because initial ownership of copyright in employment relationships is not harmonized in the EU. In several jurisdictions, such as Spain, the Netherlands, Poland, and Hungary, the employer is deemed the initial copyright owner by operation of law.[205] The language of Article 20 makes clear that the right to claim additional, proportionate remuneration is only available to authors or performers "from the party with whom they entered into a contract for the exploitation of their rights."[206] Where copyright ownership vests in the employer by statutory default, rather than being transferred through contract, it is arguable that there is no contractual baseline upon which the "best-seller" rule can operate.

If this interpretation holds, then the scope of Article 20 is significantly narrower than it might appear at first glance in employment contexts, and exclude a large segment of workers, those whose copyright is never contractually transferred, from becoming beneficiaries. This would be at odds with the broader purpose of the DSM Directive, which aims to harmonize copyright law across the EU and strengthen the position of authors and performers in the digital single market. Recital 78 clearly signals that the provision is designed to promote fairness and a more balanced bargaining environment for creators, including when creating as employees.[207] There is also a potential policy inconsistency in denying Article 20 protection to employees whose rights are appropriated by law, but allowing it where rights are

203. *Id.* at art. 19.

204. *Id.* at art. 20.

205. *See* DUSOLLIER ET AL., *supra* note 38, at 44–45; Legeza, *supra* note 32, at 12. In most Member States, however, and as discussed, the employee will by default own the copyright in employment relationships. *See supra,* Section III.

206. DSM Directive, *supra* note 10, at art. 20.

207. *Id.* at recital 78.

assigned through contract. The rationale behind the "best-seller" rule—namely, that creators should receive fair compensation if their work turns out to have significant downstream value—applies regardless of how the rights initially changed hands.

One alternative interpretation is to read Article 20 purposively, on the basis that employers are still exploiting the works produced by their employees. An employment contract could, from that perspective, still be considered an "exploitation contract" as a matter of copyright law. Even if their copyright has been assigned by statute, not contract, employees would be claiming additional remuneration "from the party [employer] with whom they entered into a contract for the exploitation of their rights."[208] Ultimately, this will be a task for the courts to decide in due course.

D. EXTRATERRITORIALITY AND CHOICE OF LAW IMPLICATIONS

The rules on fair remuneration and the "best-seller" clause in the DSM Directive are likely to have implications far beyond the borders of the EU, despite the territorial nature of copyright. Copyright arises automatically upon creation and subsists as a bundle of rights across multiple jurisdictions. Because of that, individuals working for foreign companies will be equally entitled to fair remuneration and the right to adjust their contracts with respect to their European copyright portfolio. However, these rights under Articles 18 and 20 are only triggered when the exclusive rights within the EU are implicated.[209] This analysis becomes more complex in the context of AI, which, unless embedded in physical products, is inherently digital.[210] This practically means that courts will be faced with the situation of having to put a separate price tag on the European copyright bundle, taking into account how workers' data is used within the EU specifically. That will, in turn, hinge on questions such as "how much" the data is being used within the EU and "how valuable" that use is.

To illustrate, consider a hypothetical scenario where workers are employed by a global technology company headquartered in the United States. Their data is stored on U.S. servers and used to fine-tune an LLM by teams based in the United States and Asia. The fine-tuned model is then used to deploy AI systems across global internal teams, including those in the EU. In such a case, the right to fair remuneration and contractual adjustment would apply only to the EU-specific copyright bundle across all global workers, thereby reducing any claim for

208. *Id.* at art. 20.

209. *Cf.* Rättzén, *supra* note 9, at 228–331 (discussing copyright territoriality and how the localization of cross-border infringements determine which law becomes applicable, following *lex loci protectionis*). The same general principle of copyright territoriality applies *mutatis mutandis* to Articles 18 and 20 of the DSM Directive. The DSM Directive can only regulate EU and not foreign copyright as a matter of sovereignty and public international law. *Id.* at 230–31.

210. *See* Mattias Rättzén, *Who Rules AI? The Rise of Extraterritoriality and Solving the Cross-Border AI Governance Crisis*, at 12–13 (unpublished manuscript) (on file with author).

additional remuneration. If the worker is also based in the EU, there is an argument that the price tag should be higher, especially if the downstream deployment of AI results in redundancy or job displacement within the EU. In practice, the rules on fair remuneration and "best-seller" in the DSM Directive therefore have an extraterritorial reach, and impact both workers and companies based outside the EU, but where workers' data is incorporated into AI systems deployed in the EU.

There are some open questions about whether workers based outside the EU also qualify as beneficiaries under the EU rules. The international copyright framework is based on the principle of national treatment, which provides that authors who are not nationals of the "country of origin" of the work are entitled to the same rights in a foreign country as are granted to that country's own nationals.[211] The country of origin is where the work is first published, or—if published simultaneously in multiple countries—the one offering the shortest term of protection.[212] Because nearly all countries are signatories to the Berne Convention providing this protection, it can almost always be assumed that the principle of national treatment applies. More pertinent is Article 5(2) of the Berne Convention, which provides that "the extent of protection, as well as the means of redress afforded to the author to protect his rights, shall be governed exclusively by the laws of the country where protection is claimed."[213] This long-standing rule, commonly interpreted as referring to *lex loci protectionis*, derives from the principle of territoriality.[214] In practice, it means that copyright infringement is governed by the laws of the country where the infringing act occurs, unless the same act can also be localized elsewhere.[215] Where things become less certain is in relation to contractual elements, such as licensing and assignment of rights. It is debatable whether the principle of *lex loci protectionis* extends to these areas. The reference in Article 5(2) to "extent of protection" and "means of redress" could potentially be interpreted as referring specifically to the enforcement of substantive rights, rather than to contractual arrangements.[216] However, that distinction is likely overstated in the present context. Although Articles 18–21 of the DSM Directive govern licensing and assignments, their content is substantive in nature and aimed at

211. *See* Berne Convention, *supra* note 31, at art. 5(3).

212. *Id.* at art. 5(4)(a–b).

213. *Id.* at art. 5(2).

214. *See* SAM RICKETSON & JANE GINSBURG, INTERNATIONAL COPYRIGHT AND NEIGHBOURING RIGHTS: THE BERNE CONVENTION AND BEYOND §§ 20.02–20.03 (2022); Alexander Peukert, *Territoriality and Extraterritoriality in Intellectual Property Law, in* TRANSNATIONAL LEGAL AUTHORITY IN AN AGE OF GLOBALIZATION 191–192 (Günther Handl ed., 2012); Raquel Xalabarder, *Copyright: Choice of Law and Jurisdiction in the Digital Age*, 8 ANN. SURV. INT'L & COMP. L. 79, 82–83 (2002).

215. *See* Mattias Rättzén, *Closing the Patent Loophole Across Borders*, 20 U.I.C. REV. INTELL. PROP. L. 358, 370 (2021).

216. *See* Mihály Ficsor, GUIDE TO THE COPYRIGHT AND RELATED RIGHTS TREATIES ADMINISTERED BY WIPO AND GLOSSARY OF COPYRIGHT AND RELATED RIGHTS TERMS ¶ BC-5.11, at 42 (2004); SAM RICKETSON & JANE GINSBURG, INTERNATIONAL COPYRIGHT AND RELATED RIGHTS § 6.99 (2d ed., 2006); Ted Shapiro, *Remuneration Provisions in the DSM Directive and the Audiovisual Industry in the EU: The Elusive Quest for Fairness*, 12 EUR. INTELL. PROP. REV. 42 (2020).

safeguarding authors' economic rights.[217] They do not merely set out contractual norms. These provisions also provide remedial mechanisms and fall within the "means of redress" contemplated by Article 5(2) of the Berne Convention. Accordingly, there is a strong basis for concluding that non-EU workers can also benefit from Articles 18–21 when EU copyright is at issue.

The transparency, "best-seller," and alternative dispute resolution rules are mandatory rules and cannot be contracted out of.[218] While the right to fair remuneration under Article 18 can in principle be contractually waived, Member States are free to decide otherwise.[219] Sweden, for example, has chosen to make Article 18 mandatory as well.[220] The mandatory nature of these provisions also prevents parties from circumventing them through a choice of foreign law in the contract. Recital 81 of the DSM Directive clarifies that "the parties' choice of applicable law other than that of a Member State does not prejudice the application of the provisions regarding transparency, contract adjustment mechanisms and alternative dispute resolution procedures laid down in this Directive, as implemented in the Member State of the forum."[221] This reflects Article 3(4) of the Rome I Regulation, which prevents parties from avoiding mandatory EU law where all relevant elements of the situation are located within the EU.[222] The only exception to this rule arises in cases where foreign overriding mandatory provisions apply. Specifically, Article 9(2) of the Rome I Regulation provides that "[n]othing in this Regulation shall restrict the application of the overriding mandatory provisions of the law of the forum."[223] These are provisions considered essential to safeguard a country's fundamental public interests, such as its political, social or economic organization. It is highly unlikely that Articles 19–21 of the DSM Directive would be displaced by such rules. Therefore, and to conclude, even where the contract is governed by non-EU law, the substantive rights and protections under the DSM Directive will continue to apply in full when EU copyright is involved.

VIII. IS THERE A WORKERS' DATA VALUE GAP?

A. AN ECONOMICS PRIMER OF THE VALUATION OF DATA IN AI

The right to fair remuneration and the "best-seller" rule in the EU DSM Directive opens up the possibility for workers to claim additional remuneration for the exploitation of their works. As discussed above, this includes cases where a worker's

217. DSM Directive, *supra* note 10, at arts. 18–21.
218. *Id*. at recital 81, art. 23(1).
219. *Id.* at art. 18(2).
220. Prop. 2021/22:278, *supra* note 40, at 169–70.
221. DSM Directive, *supra* note 10, at recital 81.
222. Regulation (EC) No 593/2008 of 17 June 2008 on the law applicable to contractual obligations, 2008 O.J. (L.177) 6. [hereinafter "Rome I Regulation"].
223. *Id.* at art. 9(2).

work products, if qualifying as copyright-protected works, are used as data inputs in AI systems. However, this right is only triggered when there is a clear and significant discrepancy between the economic value ultimately derived from the exploitation of those work products, including in the form of data, and the compensation originally paid. In other words, a so-called "workers' data value gap" must exist for this claim to succeed.

Pricing data is not a trivial exercise, and the economic science behind data valuation and pricing is still evolving, with little consensus.[224] The value of data is not static, and will depend on numerous factors such as its quality, quantity, relevance, and the context in which it is used.[225] The relative impact of these factors is case-specific and will vary depending on the circumstances in which the data is deployed. Data is also inherently heterogeneous, often unstructured, and difficult to reduce to standardized measurement units. Unlike physical commodities, data does not have a straightforward measurement metric, such as weight or amount, and the amount of data that is useful will depend on the application.[226] What is more, data often does not have a common market price, and there are few marketplaces from which data can be sold and bought.[227] The latter is however changing quickly, and data vendors are becoming increasingly common.

Several economic theories and methodologies have been proposed for valuing data. One of the most common is the **cost-based method**, which assesses value based on either historical costs (i.e., how much was spent to acquire the data) or replacement costs (i.e. how much it would cost to reproduce it today).[228] To assess

224. *See* Diane Coyle & Annabel Manley, *What Is the Value of Data? A Review of Empirical Methods*, BENNETT INST. FOR PUB. POL'Y, UNIV. OF CAMBRIDGE (July 2022) https://bennettschool.cam.ac.uk/wp-content/uploads/2022/07/policy-brief_what-is-the-value-of-data.pdf [https://web.archive.org/web/20251010221442/https://bennettschool.cam.ac.uk/wp-content/uploads/2022/07/policy-brief_what-is-the-value-of-data.pdf]; *see generally* Jun Hao, Zeyu Deng & Jianping Li, *The Evolution of Data Pricing: From Economics to Computational Intelligence*, 9 HELIYON (2023).

225. *See* Jian Pei, *A Survey on Data Pricing: from Economics to Data Science*, ARXIV (Sep. 9, 2020), at 32–33 (Sep. 9, 2020), https://arxiv.org/abs/2009.04462 [https://web.archive.org/web/20251010215622/https://arxiv.org/pdf/2009.04462].

226. *See* Diane Coyle & Lianna Brinded, *The Value of Data: Summary of Findings from a Literature Review*, BENNETT INST. FOR PUB. POL'Y, UNIV. OF CAMBRIDGE (Feb. 2020) at 7, https://www.bennettinstitute.cam.ac.uk/wp-content/uploads/2020/12/Value_of_data_literature_review_26_February.pdf [https://web.archive.org/web/20251010215806/https://bennettschool.cam.ac.uk/wp-content/uploads/2020/12/Value_of_data_literature_review_26_February.pdf] (discussing how other data characteristics, such as exclusivity, accuracy, completeness, consistency, interoperability, etc. matter more).

227. Santiago Andrés Azcoitia, Costas Iordanou & Nikolaos Laoutaris, *What Is the Price of Data? A Measurement Study of Commercial Data Marketplaces*, ARXIV (Oct. 25, 2021), at 1–2 https://arxiv.org/abs/2111.04427 [https://web.archive.org/web/20251005174752/https:/arxiv.org/abs/2009.04462] (finding that data products varied widely in price, from free or a few dollars to several thousand dollars).

228. FRONTIER ECONOMICS, THE VALUE OF DATA ASSETS: A REPORT FOR THE DEPARTMENT FOR DIGITAL, CULTURE, MEDIA AND SPORT 13 (Frontier Networks Ltd., 2021), https://assets.publishing.service.gov.uk/media/6399f93d8fa8f50de138f220/Frontier_Economics_-

the costs associated with data, all the costs at all stages of the data life cycle are considered, including those for generating, collecting, storing, and replacing a dataset, as well as the costs if the data were lost.[229] This approach comes from the standard practice of the System of National Accounts ("SNA"), which is the internationally recognized framework for compiling national economic statistics. The SNA endorses a "sum-of-costs" method where the value cannot be directly observed through a market transaction,[230] which is recommended for valuing items such as databases.[231] Under the SNA, relevant cost inputs include preparing data in the appropriate format, staff time spent on developing the database, capital services used in the process, and intermediate consumption costs.[232]

Despite its appeal, the cost-based method has several limitations when applied to data. It is not always obvious how data-specific costs should be isolated from other development costs, especially in AI-related projects.[233] Data costs may also vary widely depending on its role in decision-making processes or model training. Moreover, there is often no correlation between how much a dataset costs to produce and its commercial value. Data produced at low cost may have high strategic or revenue-generating potential, and vice versa.[234] As a result, cost-based methods can either significantly understate or overstate the actual economic value of data.

Income-based methods assess the expected revenue streams generated by the underlying data.[235] This approach is most helpful when revenues can be directly attributed to the data itself, but it becomes less effective when the causal link between data and income is weak or indirect.[236] A related technique, commonly used in intellectual property valuation, is the "relief from royalty" method. This estimates the cost savings a company enjoys by generating its own data instead of

---

_value_of_data_assets_-_Dec_2021.pdf [https://web.archive.org/web/20250919173626/https://assets.publishing.service.gov.uk/media/6399f93d8fa8f50de138f220/Frontier_Economics_-_value_of_data_assets_-_Dec_2021.pdf].

229. Coyle & Manley, *supra* note 224, at 5.

230. *Id.*

231. RECORDING AND VALUATION OF DATA IN NATIONAL ACCOUNTS 11 (Intersecretariat Working Group on National Accounts (United Nations), 2020), https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.20/2020/mtg1/3.4_Recording_of_Data_in_NA_Eurostat_June_2020_after_SG_comments.pdf [https://web.archive.org/web/20251010220515/https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.20/2020/mtg1/3.4_Recording_of_Data_in_NA_Eurostat_June_2020_after_SG_comments.pdf].

232. UNITED NATIONS STAT. COMM'N, SYSTEM OF NATIONAL ACCOUNTS (2008), at 10.113.

233. *See* Coyle & Manley, *supra* note 224, at 6.

234. FRONTIER ECONOMICS, *supra* note 228, at 14–15.

235. *Id.* at 8; *see also* Jenna Slotin, *What Do We Know About the Value of Data?*, GLOB. P'SHIP FOR SUSTAINABLE DEV. DATA, https://www.data4sdgs.org/sites/default/files/services_files/Value%20of%20Data%20Report_Final_compressed_0.pdf [https://web.archive.org/web/20251005174738/https:/www.data4sdgs.org/sites/default/files/services_files/Value%20of%20Data%20Report_Final_compressed_0.pdf] (last visited Aug. 17, 2025).

236. Coyle & Manley, *supra* note 224, at 8.

paying licensing fees to third parties.[237] However, this method becomes harder to apply when comparable licensing arrangements do not exist, or when the data in question is not available for purchase on the open market.

Market-based methods rely on external observables. The most important of these is the market price of the product, or users' willingness to pay for the product.[238] Where market prices are available, also known as "bottom-up," they are almost always a preferred method for valuation.[239] Yet such prices are rarely available for data, which is often created and used internally by organizations without ever being bought or sold. Where market prices are not available, it is necessary to look at what is the market price for equivalent or comparable products, or what the user would be willing to pay for that.[240] Equivalent or comparable products could be those from other companies or in related sectors. This exercise is inherently complex. Comparability is difficult to establish, and differences in use cases, quality, and sector norms can significantly impact price relevance. The drawback of this method becomes evident where data is concerned. There are so far few data marketplaces, and the price of the underlying data used in AI systems is invisible. Data is also often bundled together in larger datasets, which are sometimes made publicly available for free, making it less than ideal as a fiscal proxy.[241]

One alternative market-based method, when there are no quantifiable market prices, is using a "top-down" approach, which looks at the market value of companies instead of their goods.[242] A top-down approach begins with evaluating the entire relevant market and then estimating a particular company's market share, if there is competition. The serviceable available market will represent the total addressable market which the company can serve with its goods. Lastly, the serviceable obtainable market is the relative share of the market which the company can realistically capture, or have captured, by their goods.[243] A top-down approach is inherently speculative as it assumes that the relative market share can be attributed to the goods in question. This becomes even more difficult where data is concerned, as it often cannot be said that a company's market position can be attributed to inputs like using particular data. Success could come from a wide range of other factors, such as product execution and design, marketing, and so forth. A top-down approach would mainly be useful where the data itself is sold on the market, but less helpful when it is not. In most cases for data, when used

237. *Id.*

238. *Id.* at 9; Slotin, *supra* note 235.

239. FRONTIER ECONOMICS, *supra* note 228, at 15.

240. Slotin, *supra* note 235.

241. *See, e.g.*, COMMON CRAWL, https://commoncrawl.org/ [https://web.archive.org/web/20250920005201/https://commoncrawl.org/)], which is the largest open, free repository of web-crawled data from the internet.

242. FRONTIER ECONOMICS, *supra* note 228, at 15.

243. Daniel F. Spulber, *Finding Reasonable Royalty Damages: A Contract Approach to Patent Infringement*, 2 U. ILL. L. REV. 615, 688–89 (2019).

internally within organizations, there will simply be no "top" and no "down" as there is no market.

Impact-based methods value data based on the causal effect of its use.[244] These can be distinguished from an income-based approach in that they are concerned with the actual effect and outcome resulting from using the data. For example, using certain data may make it possible to achieve a certain technical result, which otherwise would be difficult to achieve with other data or resources. That technical result may, in turn, also yield positive financial consequences. What makes impact-based methods more difficult to apply in practice is that they require a credible counterfactual: what would have been the outcome if another set of data had been used?[245] Although that may be simulated or estimated, there will necessarily be a degree of uncertainty and a "what if" factor which can be over- or underestimated. For example, Arrieta-Ibarra et al. used counterfactual simulations to vary the amount of data which would be available for machine learning algorithms to use and compared their effect on company profits.[246] In the case of a well-known ride-share company, which used data from drivers and riders for fare estimation, analytics, and marketing, they estimated that data use could account for up to 47% of the company's annual revenue. If drivers were fully compensated for the value of the data they generate, the study suggested they could be entitled to up to $30 per day solely for their data contributions.[247]

Real options methods are concerned with not the realized value of data post-deployment, but with its potential value at the time of collection.[248] Data is often acquired without a specific use case in mind, under the assumption that it may prove valuable in the future. In this context, data functions as a strategic asset. It is useful simply by virtue of being available when future opportunities arise. Real options valuation involves simulating multiple hypothetical future scenarios in which the data might be used. The scenario value in each of these simulations is then aggregated to consider all possible outcomes.[249] For example, a company might simulate how the data could be used in different use cases, and what the outcome would be in each of these cases. Those outcomes can then be combined to better inform the decision-making process, but indirectly also the value which the data assets may possess.

There is no consensus in economic theory on what valuation method is most appropriate for data. Each method has its strengths and limitations, and it may be advisable to apply multiple methods in parallel to triangulate a more accurate valuation. Depending on the circumstances, some methods may favor data producers, while others may advantage data users. Relying solely on one

244. Slotin, *supra* note 235; Coyle & Manley, *supra* note 224, at 15.

245. Coyle & Manley, *supra* note 224, at 16.

246. *Id.* at 15 (citing Imanol Arrieta Ibarra et al., *An Empirical Study of the Value of Data* (2020) (unpublished manuscript) (on file with Mimeo)).

247. *Id.*

248. *Id.* at 18.

249. *Id.* at 19.

methodology may lead to distorted or incomplete assessments. The analysis will also necessarily be context-specific. Data which improves the performance of a model in one scenario may not do so in another, meaning that the purpose behind using the data should be a key factor to consider in valuation.

AI applications and machine learning algorithms rely heavily on data for their training. Yet the data's value varies considerably depending on how it is used and the type of AI technology involved. In supervised learning, where labeled datasets train models to make predictions, data quality and relevance have a direct and measurable effect on accuracy.[250] By contrast, unsupervised learning systems, which identify patterns in unlabeled data, derive value primarily from the volume and diversity of datasets.[251] Large and varied datasets can be especially valuable in deep learning, which typically requires vast amounts of data.[252] Fine-tuning pre-trained models is another area where data can hold significant value. In such cases, smaller, domain-specific datasets are often used to adapt a general-purpose model to a more targeted task.[253] Data quality and domain relevance becomes paramount for this purpose. A similar principle applies to retrieval-augmented generation ("RAG") systems, which combine pre-trained language models with external knowledge retrieval. The value of data in RAG systems lies in its relevance to the retrieval task, meaning the more specific and relevant it is to the task, the more valuable it will be.

Lastly, the deployment context matters. Data used for AI deployed in high-risk applications, such as autonomous vehicles or medical diagnostics, might be valued more than data used for low-risk applications.[254] Some data will also be more difficult to substitute than others, particularly where it derives exclusively from internal sources.[255] Additionally, the value of data can differ depending on whether the AI technology is designed for a one-time task or continuous learning. In systems that continuously update and improve over time, such as recommendation engines, the ongoing collection of new data to refine the model will be particularly valuable.

---

250. *See* Saket Srivastava et al., *Impact of Data Quality on Supervised Machine Learning: Case Study on Drilling Vibrations*, 219 J. PETROLEUM SCI. & ENG'G 111058 (2022).

251. *See* Shayne Longpre et al., *A Pretrainer's Guide to Training Data: Measuring the Effects of Data Age, Domain Coverage, Quality, & Toxicity*, ARXIV, No. 2305.13169, at 3 (Nov. 13, 2023), https://arxiv.org/abs/2305.13169 [https://web.archive.org/web/20251005174847/https:/arxiv.org/abs/2305.13169].

252. Michael R. Douglas, *Large Language Models*, ARXIV, No. 2307.05782, at 5–7 (Oct. 6, 2023), https://arxiv.org/abs/2307.05782 [https://web.archive.org/web/20251005174955/https:/arxiv.org/abs/2307.05782]; *see also* Google LLC, *Comment Letter on Artificial Intelligence and Copyright*, 88 FED. REG. 59942, at 3–4 (Oct. 30, 2023).

253. Kunpeng Guo et al., *Fine-tuning Strategies for Domain Specific Question Answering Under Low Annotation Budget Constraints*, ARXIV, No. 2401.09168, at 3 (Jan. 14, 2024), https://arxiv.org/abs/2401.09168 [https://web.archive.org/web/20251010222726/https://arxiv.org/pdf/2401.09168].

254. *Cf.* Coyle & Brinded, *supra* note 226, at 6–7 (discussing how liability issues relating to data could reduce the value of data); *see also* FRONTIER ECONOMICS, *supra* note 228, at 65.

255. *See* Coyle & Brinded, *supra* note 226, at 7.

This contrasts with AI systems designed for static tasks, where the initial dataset might represent most of the data's value.

B. WHAT MAKES THE VALUATION OF WORKERS' DATA UNIQUE

Workers' data differs from other data used in AI because it is generated in the course of employment. This exclusivity makes workers' data unique, context-specific, and generally irreplaceable. As a result, there are typically no external markets or substitutes. It is also this lack of substitutability that makes workers' data particularly valuable to organizations. Moreover, unless workers' work products have been shared with third parties or made public, competitors are unlikely to have access to equivalent datasets.

Market prices usually provide the most reliable valuation benchmark for data, but when no such market exists for workers' data, other approaches such as **cost-based methods** may be used. These focus on the costs associated with data generation that can be attributed to the worker, including salary, bonuses, or other forms of compensation paid.[256] Broader overhead costs, such as those for systems, tools, or infrastructure used to capture, store, and process the data, should be excluded from the calculations. But even so, wages may still misstate the economic value, as they often reflect factors unrelated to the worth of the data produced. In some roles, data creation is incidental, while in others it represents the employee's main contribution. Where the latter applies, the data's value may already be fully or partly captured in pay. In other cases, workers may simply receive high or low compensation for reasons unrelated to the value of the work products they generate.

What complicates the economic analysis of workers' data is that workers may have already received fair compensation for the work products they generated. The challenge arises when those same work products, once transformed into data, are repurposed for new applications that create additional value and revenue streams for the organization. Traditional cost-based valuation methods, which focus on the initial costs of producing the data, such as employee wages, may not fully capture the economic value generated by the data over time.[257] Workers' data can be used to develop entirely new applications in AI and machine learning, far exceeding the initial value of individual work products. This suggests that cost-based methods, while useful as a baseline, are insufficient on their own, and that complementary methods, such as **income-based valuation**, are necessary to account for the downstream value derived from data reuse. AI systems which are fine-tuned or developed internally within the organization could also be used externally, in which case these will have their own attributable revenue. The income generated from selling or licensing these AI systems should be considered when valuing workers'

256. *Cf.* Coyle & Manley, *supra* note 224, at 5 (discussing the valuation principles in respect of data generally).

257. *See id.* at 7 (discussing how data may change its value over time).

data. If the AI system is only used internally, the focus shifts to evaluating internal efficiencies and cost savings. Here, the value of workers' data could be measured by how much it reduces operational costs or improves productivity or decision-making. For example, if an AI system developed using workers' data significantly reduces the time required for a particular task, or means that fewer workers are needed to perform the same task, the cost savings could be substantial, which should be a factor when valuing the data.

Impact-based methods can also provide valuable insights. Workers' data may drive measurable outcomes, such as process optimization, that lead to tangible financial returns. In product development, workers' data could also contribute to new improvements, or in exceptional cases lead to new inventions. However, a key challenge of the impact-based approach lies in establishing the counterfactual: what would have occurred in the absence of using the workers' data? If similar outcomes could have been achieved using alternative data sources, the unique value of the workers' data is diminished from an impact-based perspective. Finally, where workers' data is collected with a view to potential future use, but no concrete use case yet exists, **real options methods** can be helpful. These methods rely on simulating hypothetical future scenarios, which necessarily involves a degree of speculation and often relies on sophisticated modelling techniques to generate meaningful valuation results.

Apart from choosing valuation methods, another important factor to consider will be who is contributing to the dataset and who is using the dataset. In scenarios where multiple data producers provide data to a single user, it will be difficult to measure the value of individual data contributions. In these cases, pricing strategies instead calculate the marginal contribution of data. The so-called Shapley value, which is a concept originating from cooperative game theory, is often used for such data valuation exercises.[258] The Shapley value is defined using four axioms, which should each be satisfied: efficiency, symmetry, the zero element, and additivity. The exact calculation of the Shapley value requires evaluating all possible combinations of data points, which is computationally expensive and infeasible for large datasets. More specifically, the Shapley value for an agent *i* is given by:

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} (v(S \cup \{i\}) - v(S))$$

where:

N is the set of all agents (data contributors).

S is a subset of *N* not including agent *i*.

258. *See* Jimin Xu et al., *Data-Driven Learning for Data Rights, Data Pricing, and Privacy Computing*, 25 ENG'G 66, 70 (2023) (citing Lloyd S. Shapley, *A Value for n-Person Games, in* CONTRIBUTIONS TO THE THEORY OF GAMES (Harold W. Kuhn & Albert W. Tucker eds., 2016)).

$v(S)$ is the value of the coalition S.

$v(S \cup \{ i \})$ is the value of the coalition S plus agent i.

In a cooperative game, where multiple agents (or, in this case, data contributors) contribute to a collective outcome, the Shapley value determines how to distribute the total payoff among the participants based on their individual contributions. Efficiency ensures that the total value generated by the coalition is fully distributed among the participants. Symmetry ensures that if two agents contribute equally, they receive equal shares. The zero element principle excludes any agent that adds no value from receiving any payoff, and additivity allows for the combined value of agents to be considered across different coalition games. As Xu et al. have recognized, however, calculating the Shapley value presents significant challenges in the context of AI and machine learning.[259] The exact calculation requires evaluating all possible combinations of data points, which is computationally expensive and infeasible for large datasets. Several other methods have been developed to address this. Ghorbani & Zou introduced Monte Carlo and gradient-based techniques to approximate Shapley values efficiently in supervised machine learning contexts.[260] Jia et al. also developed a method for calculating exact Shapley values in K-nearest neighbors ("KNN") models with improved computational efficiency.[261] In another neighboring method, Xu et al. have proposed to determine the value of data by its intrinsic nature, where the volume of the dataset is used as a coalition revenue function.[262] That is to say, the volume of the data, both in total and relative from each data contributor, is used to proxy the value.

The real challenge in assessing the value of individual data contributors does not lie in calculating the relative distribution of the value among them, but rather in accurately estimating each contributor's marginal contribution. To determine the marginal contribution of an individual data contributor, it is necessary to compare the model's performance with and without that contributor's data.[263] This involves creating counterfactual scenarios where the data from the contributor is either included or excluded.[264] This exercise is inherently complex, especially in the context of AI and machine learning. For example, if a dataset contains data from contributors A, B, and C, it is necessary to train the model multiple times: once with all data, and then once without each contributor's data (i.e., without A, without B, and without C, respectively). Only then is it possible to more accurately assess what

259. *Id*.

260. Amirata Ghorbani & James Zou, *Data Shapley: Equitable Valuation of Data for Machine Learning*, 97 PROCS. OF THE 36TH INT'L CONF. ON MACH. LEARNING (ICML) 2242, 2242 (2019).

261. Jia Ruoxi et al., *Efficient Task-Specific Data Valuation for Nearest Neighbor Algorithms*, 12(11) PROC. VLDB ENDOW. 1610, 1613 (2019).

262. Xinyi Xu et al., *Validation Free and Replication Robust Volume-Based Data Valuation,* 2021 PROCS. OF 35TH CONF. ON NEURAL INFO. PROCESSING SYS. 10837, 10838.

263. Huaiguang Cai, *CHG Shapley: Efficient Data Valuation and Selection Towards Trustworthy Machine Learning*, ARXIV, No. 2406.11730, at 2–3 (June 17, 2024), https://arxiv.org/abs/2406.11730 [https://web.archive.org/web/20251010223213/https://arxiv.org/pdf/2406.11730].

264. Ghorbani & Zou, *supra* note 260, at 2244.

the relative impact of an individual contributor's data is on model performance, and ultimately its marginal contribution and economic value.

These counterfactual analyses are far from practical in legal settings. Conducting them with precision would likely require significant economic modeling and scientific expert evidence. A more pragmatic approach in the context of copyright and data valuation is to rely on **approximation methods**. Monte Carlo simulations involve sampling a subset of all possible coalitions of data contributors and averaging the marginal contributions, but since then even more simplified techniques have been proposed in economic literature.[265] These new techniques rely on only small sample sizes, while still producing reliable approximations. It may also be useful to draw analogies from how marginal contribution issues have been handled in other domains of law. For example, in FRAND ("Fair, Reasonable, and Non-Discriminatory") patent litigation, courts have long grappled with the problem of estimating the relative value of individual patents within large patent portfolios. Some of the legal and economic theories developed in that context may offer useful insights for addressing similar valuation challenges in the context of workers' data. These parallels are explored further in subsequent sections.[266]

C. WHETHER WORKERS SHOULD BE COMPENSATED FOR THE USE OF THEIR DATA

Whether there is a workers' data value gap will depend heavily on the circumstances.[267] On one hand, any compensation paid to workers already is likely to account for the fact that work products, at least to some extent, may contain data that could be used for AI or automation. Producing such outputs is often a worker's primary contractual duty, and once the copyright in those works vests in the employer or contractee, the worker is presumed to have received fair remuneration under a cost-based analysis. The employer or contractee is then free to exploit the data embedded in those works without owing further obligations. On the other hand, Article 20 of the DSM Directive unsettles this premise by introducing a contract adjustment mechanism.[268] This shifts the focus away from the formal bargain at the time of contracting to whether remuneration remains proportionate to the actual or potential economic value later extracted. A cost-

265. Lauren Watson et al., *Accelerated Shapley Value Approximation for Data Evaluation*, ARXIV, No. 2311.05346, at 1 (Nov. 9, 2023), https://arxiv.org/abs/2311.05346 [https://web.archive.org/web/20251005175208/https:/arxiv.org/abs/2311.05346].

266. *See infra* Section IX.D.8.

267. *Cf.* David Nguyen & Marta Paczos, *Measuring the Economic Value of Data and Cross-Border Data Flows: A Business Perspective*, OECD DIGIT. ECON. PAPERS (Aug. 26, 2020), at 32-36, https://www.oecd.org/content/dam/oecd/en/publications/reports/2020/08/measuring-the-economic-value-of-data-and-cross-border-data-flows_219e1b8b/6345995e-en.pdf [https://web.archive.org/web/20250919212326/https://www.oecd.org/content/dam/oecd/en/publications/reports/2020/08/measuring-the-economic-value-of-data-and-cross-border-data-flows_219e1b8b/6345995e-en.pdf] (discussing the relative and circumstantial nature of economic analysis in valuing data).

268. DSM Directive, *supra* note 10, at art. 20.

based method for calculating the economic value of copyrighted works and their data may be inadequate on its own, in which case it opens up the possibility of valuing workers' data using income-based, impact-based, and real options methods.

The challenge is that economic theories of valuation need to align with copyright policy. This is essential not only because workers may qualify as authors, but also because any claim for additional compensation is grounded in copyright law, not contract. Copyright is justified on the premise that creative works exhibit quasi–public good characteristics, being non-rivalrous and partly non-excludable.[269] Its non-rival nature lies in that the consumption of the good by one person does not reduce the supply available for consumption by another.[270] Its non-excludable nature refers to that the use by one person neither prevents the access by other people simultaneously.[271] Copyright seeks to address the market failure that occurs when such works are underproduced or underutilized due to the risk of free-riding.[272] By conferring a statutory monopoly on rightsholders, copyright enables them to exclude others and charge prices above marginal cost, thus encouraging investment in creative production.[273]

Applying these principles to workers' data is not straightforward. Workers' data is a byproduct of work products, for which wages already serve as the incentive mechanism. Nor is it necessarily non-excludable, since companies routinely deploy technological and contractual means to restrict access.[274] Furthermore, the potential for market failure is not obvious in employment or contractor settings. If someone has accepted a job and gets paid for it, they will execute the role, which will result in the creation of work products. Companies may argue that rewarding workers for their data is actually a free-riding problem in itself. They will have invested in the personnel, infrastructure, and systems that enable the collection and exploitation of workers' data. From that point of view, it is therefore companies, not workers, that must be incentivized to exploit workers' data if we are to avoid market failure. The broader societal benefits of extending statutory protection or compensation rights to workers' data are also uncertain. While the use of workers' data may lead to the development of more effective AI systems, which could benefit society as a whole, it is not clear how these benefits could or should be fairly distributed among the many workers whose contributions form the underlying data.

---

269. *See* William M. Landes & Richard A. Posner, *An Economic Analysis of Copyright Law*, 18 J. LEGAL STUD. 325, 326 (1989); Mark A. Lemley, *The Economics of Improvement in Intellectual Property Law*, 75 TEX. L. REV. 989, 994–99 (1997); Stephen Breyer, *The Uneasy Case for Copyright: A Study of Copyright in Books, Photocopies, and Computer Programs*, 84 HARV. L. REV. 281 (1970).

270. *See* Christopher S. Yoo, *Copyright and Public Good Economics: A Misunderstood Relation*, 155 U. PA. L. REV. 635, 637 (2007).

271. *Id.*

272. *Id.* at 642–45.

273. *Id.* at 646–48.

274. *See supra* Section II (discussing how workers' data is often deployed exclusively within organizations).

Workers may nonetheless contend that they are entitled to a share of the additional value their data generates, particularly when that value is significant and when the data is redeployed in ways not contemplated at the time of contracting. This argument is most compelling when workers are displaced by automation trained on their own outputs.[275] Extracting workers' data to automate work tasks and replacing the authors who contributed to that data is, in a way, a market failure in itself. This tension becomes even more pronounced when viewed through the lens of copyright. Courts in several jurisdictions have confirmed that AI-generated content cannot attract copyright protection absent human authorship.[276] If automation increasingly replaces human creators, the volume of copyrightable works may diminish. While some may argue that copyright law will become obsolete in the age of AI, if human creativity is no longer central to the production process,[277] others may view this development as eroding the incentive structure on which copyright is premised. In the long term, there is a risk of systemic decline in creative labor markets. A related, even if extreme, market failure is that companies exploiting workers' data to replace their own workforce could eventually face a situation where fewer workers are willing to work for them.

The "best-seller" right to contract adjustment in Article 20 of the DSM Directive reflects several of these broader concerns.[278] It is a statutory mechanism that does not sit comfortably within the established economic theories of copyright. If authors have already signed contracts, been paid, and delivered their work, why should the law intervene? The European Commission, in its impact assessment preceding the first draft of the DSM Directive, explained that "[t]he contract adjustment mechanism could remedy those cases in which a lump-sum/buy-out deal turns out to be unfair, and it also addresses outright unbalanced deals as well as changed circumstances. The mechanism would reinforce creators' bargaining

275. *Id.*

276. *See* Thaler v. Perlmutter, 687 F. Supp. 3d 140, 146–150 (D.D.C. Aug. 18, 2023) (affirming the U.S. Copyright Office's position that AI-generated artwork is not eligible for copyright protection under U.S. law and explaining that human authorship is a "bedrock requirement of copyright"); Naruto v. Slater, 888 F.3d 418, 425–426 (9th Cir. 2018) (holding that only works created by a human can be copyrighted under U.S. copyright law); Městský soud v Praze 13.10.2023 [Decision of the Mun. Ct. Prague of 2023] č. j. 10 C 13/2023-16 (Czech.) (holding that an image created by Open AI's Dall-E was not protectable by copyright). Although there is no similar case law yet in the UK, its copyright law extends the concept of authorship to "the person by whom the arrangements necessary for the creation of the work are undertaken." Copyright, Designs and Patents Act 1988, c.48, § 9(3) (UK). It is an open question, however, if this means that AI-generated works could be protected by copyright where there is little to no human supervision, and, if so, who should be the author. In China, the Beijing Internet Court ruled in November 2023 that an AI-generated image generated using more than 150 prompts and negative prompts, and setting various parameters, could be awarded copyright protection as it was the direct result of a human's intelligent input and individual expression. Li Moumou Su Liu Moumou Qinhai Zuopin Shumingquan Xinxi Wangluo Chuanboquan Jiufen An (李某某诉刘某某侵害作品署名权、信息网络传播权纠纷案) [Li v. Liu], Beijing Internet. Ct. No. 11279, Nov. 27, 2023 (China).

277. *See* Mark A. Lemley, *How Generative AI Turns Copyright Law Upside Down*, 25 COLUM. SCI. & TECH L.R. 21, 40 (2024).

278. DSM Directive, *supra* note 10, at art. 20.

position."[279] While one might argue that market failure may arise over time if there is a persistent mismatch between the value of content produced and the value extracted from it, that was not the Commission's primary rationale. Recital 78 confirms this.[280] Instead, Article 20 is, at its core, a protectionist measure designed to promote fairness and rebalance contractual power between creators and exploiters. The same fairness concerns extend to all exploitation contracts, including both employment and contractor settings.

Ultimately, whether workers should be compensated for the use of their data depends on whether a clear and significant imbalance exists between what companies gain (work products and their derivative use as data) and what workers receive (their agreed pay and benefits). This is not a simple question to answer, because valuing data, let alone workers' data, is not simple. What the preceding discussion does suggest, however, is that a cost-based valuation method is often insufficient on its own. To capture the full value of workers' data, alternative valuation metrics may be necessary.

Both income-based and impact-based methods suggest that workers' data can yield substantial economic value, particularly where specialized datasets are required for AI systems and cannot be obtained elsewhere.[281] If those systems are subsequently deployed to displace the workers, whose data trained them, the resulting cost savings are direct evidence of the added value. That value may persist indefinitely. Once developed, AI systems can continue generating returns long after the workers have left, while the displaced workforce bears the cost of job loss without sharing in ongoing gains. Workers producing copyrighted works as part of their jobs are arguably in a more vulnerable position than ever. Altogether, this implies that in some cases a workers' data value gap could exist, while in others it does not.

To illustrate this, consider the following scenario: an illustrator is employed to create original illustrations. These illustrations are used as training data to fine-tune an AI system capable of generating illustrations that closely resemble the original works, or which can be adapted across various use cases. If the illustrator is subsequently made redundant, the lost economic value could potentially be substantial. In such a case, the "best-seller" rule of Article 20 of the DSM Directive becomes relevant, as it applies to all forms of economic value: actual and potential, past, present, and future.[282] But how should we value an AI system that can indefinitely generate products that take advantage of a human worker's output? The simple example below can help clarify this.

Facts:

279. *Commission Impact Assessment*, *supra* note 146, at 187.

280. DSM Directive, *supra* note 10, at recital 78.

281. FRONTIER ECONOMICS, *supra* note 228, at 27, 65 (highlighting the exclusiveness or scarcity of data as an important factor when valuing data, among other factors).

282. DSM Directive, *supra* note 10, at art. 20.

Assume the illustrator working full-time is paid an annual salary of €50,000 for creating original illustrations, which would average to around 200 works per year. Now, suppose that the fine-tuned AI model using the illustrator's data can generate illustrations of similar quality, requiring limited human supervision from someone working part-time, at a rate of 2,000 illustrations per year. The part-time human supervisor is paid an annual salary of €10,000 per year. The AI model costs €100,000 to develop, and costs €10,000 per year to maintain and update. Each illustration can be sold or used by the company to generate €600 in revenue.

Scenario A: Full-time human illustrator working over 10 years:

Total human-produced revenue:
200 works per year x €600 per work = €120,000 x 10 years = €1,200,000

Total human-related costs:
€50,000 salary x 10 years = €500,000

Net value (total revenue - total costs):
€1,200,000 - €500,000 = €700,000

Scenario B: Part-time human illustrator in combination with AI over 10 years:

Total human-produced and AI-generated revenue:
2,000 works per year x €600 per work = €1,200,000 x 10 years = €12,000,000

Total human-related costs:
€10,000 salary x 10 years = €100,000

Total AI-related costs:
€100,000 development cost + €10,000 annual maintenance cost x 10 years = €200,000

Net value (total revenue - total costs):
€12,000,000 - €300,000 = €11,700,000

Comparison of annual net value in Scenario B from Scenario A:

Annual net value in Scenario A: €120,000 - €50,000 = €70,000
Annual net value in Scenario B: €1,200,000 - €110,000 = €1,090,000
Increased annual net value from Scenario A to B: €1,090,000 - €70,000 = €1,020,000

Comparison of 10-year net value in Scenario B from Scenario A:

10-year net value in Scenario A: €700,000
10-year net value in Scenario B: €11,700,000
Increased 10-year net value from Scenario A to B: €11,000,000

In Scenario A, the full-time illustrator's cost-to-value ratio over a 10-year period would be approximately 42% (€500,000 / €1,200,000), and the entire €700,000 net value would be derived from human effort alone. In Scenario B, the AI combined with only using a part-time illustrator will allow the production of significantly more

illustrations per year, resulting in a 10-year net value increase which is more than 16 times higher than in Scenario A (€11,700,000 / €700,000). Meanwhile, the total cost-to-value ratio over a 10-year period would be as low as 2.5% (€300,000 / €12,000,000). The AI, leveraging the data derived from the illustrator's original works, therefore contributes to a massive increase in productivity and revenue. That together, with having significantly less costs spanning over a longer period of time, means that the economic output and value is significantly higher in Scenario B than Scenario A. In actual terms, the company's illustration practice is more than 16 times (!) more profitable in Scenario B, thanks to AI.

This strongly suggests that there could be a workers' data value gap in certain circumstances, and that the worker may be entitled to adjust his or her contract, now terminated, and claim further remuneration under Article 20 of the DSM Directive. The extent of that remuneration would, however and again, depend on what assumptions were made between the parties at the time of contracting. It would also largely depend on whether only the workers' work products were used as data for developing the AI, or if other data sources were used as well. How important the data content is for the AI outcome is another important factor to consider. The relative weight of these factors, and what economic proxies should be used to estimate the relative share of the economic value owed to the worker, will need to be considered in future litigation. Following sections discuss these and related questions in greater detail.

IX. BRIDGING THE WORKERS' DATA VALUE GAP

A. EXISTING CONTRACT AND EMPLOYMENT REMEDIES INADEQUATELY PROTECT WORKERS' DATA RIGHTS

The current legal framework falls markedly short in safeguarding workers' rights over the use of their data by employers or contractees. Much of the work produced by workers, and data contained therein, will be protected by copyright. As discussed above, jurisdictions will either assign the copyright to workers as first authors, or automatically assign it to their employers and in some cases contractees.[283] In practice, however, copyright ownership is routinely contractually assigned to employers and contractees in many industries.[284] Employees and independent contractors are also routinely expected to provide their consent for the processing of any personal data contained in their work products.[285] The freedom of contract trumps, leaving workers with few remedies left in their toolbox in most jurisdictions.

Judicial intervention in these cases is rare. Courts are generally reluctant to set aside contracts or particular clauses on account of unconscionability, which sets a

283. *See supra* Section III.
284. *See Id.*
285. *See supra* Section IV.A.

high legal bar to meet. Unconscionability boils down to the notion that contracts should not be unfairly exploitative or oppressive to one party, and exploited by the other.[286] There are legitimate reasons for employers and contractees to maintain ownership of their workers' data, and if it would ever be considered unconscionable it would be in exceptional circumstances only.[287] For similar reasons, it would be difficult to mount a case that employers or contractees are unjustly enriched by using workers' data which they lawfully own.[288] Moreover, both unconscionability and unjust enrichment are doctrines of the common law and are largely absent in civil law systems.

One exception can be found in the Nordic countries, where courts have a limited power to void unreasonable contract terms under general principles of fairness.[289] Even so, the legal threshold of unreasonableness to meet in such cases is still high, particularly in commercial settings, and courts are still generally reluctant to adjust the terms or conditions for commercial contracts. This concern was recognized during Sweden's implementation of the DSM Directive. The government concluded that the contract adjustment mechanism under Article 20 offered a lower threshold than the doctrine of unreasonableness under general contract law.[290] By contrast, most continental European legal systems lack any equivalent doctrine applicable to business-to-business relationships, and their courts will typically uphold the sanctity of contract unless clearly overridden by statute. Employment law and AI regulations similarly fail to address the issue of using workers' data for automation, and ultimately redundancy, purposes.[291]

Although the recently adopted EU Data Act includes provisions that might conceivably be cited to support a claim concerning workers' data, the validity of such an interpretation remains uncertain. The law is primarily geared toward Internet-of-Things data sharing, and its structure and underlying intent do not plainly accommodate employment relationships or internal organizational datasets. A literalist reading may allow for some worker-related implications, but these are neither explicit nor easily actionable in practice.[292]

The net result is a substantial gap in existing legal protections for workers. Contract, employment, and data law all fall short of offering redress when workers' data is used for automation purposes, even if done to the detriment of those who created the data in the first place. Although many companies will be using workers' data for legitimate purposes, the current legal framework, or lack thereof, opens up the possibility of a "shadow" hiring culture, where skilled workers are hired for an initial period of time to accumulate enough data, but once accumulated, face reduced working hours or redundancies as automation replaces more tasks. It is

286. *See supra* Section IV.B.
287. *See supra* Section V.
288. *See supra* Section IV.C.
289. *See* CLAES-ROBERT VON POST, *supra* note 71.
290. Prop. 2021/22:278, *supra* note 40, at 174–75.
291. *See supra* Sections IV.D & E.
292. *See supra* Section IV.B.

ultimately a policy, not legal, question whether or not this form of practice should be accepted.

B. WHAT LESSONS CAN BE LEARNED FROM PATENT LAW FOR COMPENSATION FOR EMPLOYEE-INVENTIONS?

The question of compensating workers for their creations finds a useful parallel in patent law. Employees in design and research capacities often create inventions that generate substantial value for their employers. Recognizing this, many jurisdictions have introduced statutory mechanisms allowing employees to claim additional compensation for inventions created in the course of their employment, in addition to any contractual remuneration, such as salary or bonuses.

In the UK, employees who make inventions are entitled to a "fair share" of the reward to the employer.[293] The right to compensation differs depending on if the invention initially belongs to the employer or the employee. If the employee invents something that belongs to the employer from the outset, typically due to contractual terms, then the employee may apply for compensation only if the invention has conferred an "outstanding benefit" on the employer.[294] If, however, the invention initially belongs to the employee but is later assigned or licensed to the employer, the employee may claim compensation if the benefit to them is "inadequate" compared to that received by the employer.[295] In such cases, there is no requirement to prove that the invention was of "outstanding benefit." The amount of the "fair share" compensation is determined to account for the benefit which the employer has derived, or may reasonably be expected to derive, having considered all the circumstances.[296] Among the factors considered include the employee's duties and remuneration, the skill and effort contributed by the employee and others, and the employer's support, such as facilities, guidance, and commercial or managerial expertise.[297]

Similar laws that grant employee-inventors a right to claim additional compensation for their inventions have been introduced in other European countries. In Germany, the employee will have initial ownership of the invention. Employees are obliged to offer the right to use "service inventions" to their employers,[298] for which employers have a duty to pay reasonable compensation.[299] Service inventions are those which have resulted from the employee's tasks in the

293. Patents Act 1977, § 41 (UK).
294. *Id.* at § 40(1).
295. *Id.* at § 40(2).
296. *Id.* at § 41(1).
297. *Id.* at § 41(4).
298. Gesetz über Arbeitnehmererfindungen [ArbNErfG] [German Employees' Inventions Act] July 25, 1957, BUNDESGESETZBLATT TEIL I at § 6(1). For an English translation, *see* https://www.dpma.de/docs/dpma/schiedsstelle/employee_inventions_act.pdf [https://web.archive.org/web/20251010224457/https://www.dpma.de/%20docs/dpma/schiedsstelle/employee_inventions_act.pdf].
299. *Id.* at § 9(1).

work, or which are essentially based upon the experience or activities of the employer.[300] Multiple factors need to be taken into account when considering the amount of compensation, including the commercial applicability of the service invention, the duties and position of the employee, and the employer's contribution to the invention.[301]

Such laws also exist in France,[302] Sweden,[303] the Netherlands,[304] and many other countries in Europe and in other parts of the world. The situation in the United States differs significantly. Under U.S. law, the default rule is that inventors retain ownership of their inventions, unless assigned to their employers.[305] However, American courts have granted employers so-called "shop rights," which grant the employer a non-exclusive, royalty-free license to use any inventions created by an employee during the course of their employment.[306] Employee-inventors are not entitled under statute to any compensation for such licenses to their employers, which must be contractually agreed between the parties.[307] However, because the license is non-exclusive, employee-inventors are, unless otherwise agreed, free to commercialize their inventions outside the employment context to obtain additional compensation.

Employees' right to compensation for employee-inventions has been justified on the basis that such inventions may generate substantial value for their employers.[308] These inventions, arising from the employee's labor, skills, and use of employer resources, contribute directly to corporate success. The rationale behind compensating for those inventive efforts is rooted in equity, as market-based arrangements may be inadequate. In the UK, before the Patents Act 1977, inventions created in the course of employment were owned by the employer, with employees having no right to additional compensation. The Report of the Committee to Examine the Patent System and Patent Law criticized this "master and servant" approach as inequitable, noting that it was not in line with "modern views on industrial relations."[309] In particular, it was considered that the old regime was unfair when employees produced exceptional inventions that led to substantial profits to their employers, for which they obtained no share.[310] Employees are typically in a worse bargaining position than their employers and would, in the

300. *Id.* at § 4(2).

301. *Id.* at § 9(2).

302. Code de la propriété intellectuelle [Intellectual Property Code] art. L. 611–7 (Fr.).

303. Lag om rätten till arbetstagares uppfinningar [Act (1949:345) on the Right to Employee's Inventions] (Svensk författningssamling [SFS] 1949:345) (Swed.).

304. Rijksoctrooiwet 1995 [Dutch Patent Act], Stb. 1995, 109, art. 12(1) (Neth.).

305. *See* Banks v. Unisys Corp., 228 F.3d 1357, 1359 (Fed. Cir. 2000).

306. *See* United States v. Dubilier Condenser Corp., 289 U.S. 178, 187 (1933); Mechmetals Corp. v. Telex Comput. Prods., Inc., 709 F. 2d 1287, 1291 (9th Cir. 1983).

307. *See Dubilier Condenser Corp.*, 289 U.S. 178, 188–89 (1933).

308. *See* Hanns Ullrich, *Harmonization of Employee Invention Laws: The Black Hole of the EU's Innovation Policy*, MAX PLANCK INST. FOR INNOVATION & COMPETITION RSCH. PAPER SERIES (2022), at 25–29.

309. *See* Report of the Committee to Examine the Patent System and Patent Law, Cmnd. 4407 (1970) (U.K.), at ¶ 459 (cited in Kelly & Anor v. GE Healthcare Ltd [2009] EWHC 181 (UK)).

310. *Id.*

absence of statutory intervention, likely only be able to secure significantly less compensation. Statutory compensation schemes prevent that from happening. Adequate compensation for employee-inventions also follows the public policy rationale of the patent system to reward and incentivize inventions.[311]

Patent law thus ensures compensation for intellectual contributions that create value beyond ordinary remuneration. There are several parallels which can be drawn from that to using workers' data for AI development and deployment. Like employee-inventors, workers generate valuable data through their expertise and daily tasks, which will possess increased value in the age of AI. When companies use such data to train AI systems that reduce costs or displace workers, they effectively monetize employee labor without additional reward. Just as patent laws have evolved to protect employee-inventors and ensure they receive a fair share of the benefits their inventions generate, there is a case for adopting similar protection in the case of workers' data in copyright law.

The analogy, however, has its limits. Patentable inventions are exceptional, novel, and non-obvious creations, whereas much of workers' data are generated from routine tasks that lack such qualities. There is no policy need to incentivize routine work, for which employees are already remunerated. Rather, the policy concern lies in the employer's computational use of those routine outputs to extract disproportionate value. That value is not adequately returned to workers, and in fact, the opposite is true if the AI systems using workers' data are used to make those same workers redundant. The value of such data can also be long-term, even potentially indefinite in time. Thus, while patent law provides a helpful starting point for framing equitable compensation in the data economy, there are important distinctions to be made.

C. TOWARDS A CONTRACT ADJUSTMENT MECHANISM FOR THE USE OF WORKERS' DATA

The current legal framework concerning the use of workers' data offers limited protection against potential exploitation in an era increasingly dominated by AI technologies. At present, the most promising legal remedy available in the EU is found in the DSM Directive, which provides a right to fair remuneration for the exploitation of copyrighted works, as well as a contract adjustment mechanism, known as the "best-seller rule."[312] Originally intended to address imbalances in the remuneration of authors in the digital distribution of works (e.g. audiovisual content), the DSM Directive was not, at the time of its drafting, designed with AI or data-driven automation in mind. Nevertheless, its provisions are broadly worded

311. Ullrich, *supra* note 308, at 28–29.
312. DSM Directive, *supra* note 10 at art. 20.

and, on their face, apply to all forms of copyright licensing and assignment, including those in employer-employee or contractee-contractor relationships.[313] This interpretation is supported by government statements in Member States such as Sweden, confirming that the Directive's application to such relationships was discussed during its transposition into national law.[314] Whether the framework extends to situations where workers' copyrighted work products are used as data to develop AI systems is ultimately a question for judicial interpretation, which is explored in more detail below.

Importantly, the right to fair remuneration can be waived by contract.[315] Whether such a waiver exists in a given case will often depend on how the contract is drafted and interpreted. If compensation for the use of workers' data was never discussed or expressly excluded at the time of contracting, the legal position may be unclear. Of course, this can be easily rectified by expressly including, or excluding, the right to fair remuneration in the contract; no doubt an important lesson to be learned for all drafting parties.

What cannot be waived, not even by choosing the governing law of the contract, is the "best-seller" rule.[316] The "best-seller" rule allows authors to claim additional, appropriate, and fair remuneration when the actual or potential economic value of their work turns out to be significantly higher than originally estimated, regardless of whether they have assigned their rights.[317] The main policy goal is to ensure that authors receive a fair share of the full economic value generated from the exploitation of their works.[318] If the remuneration initially agreed proves to be disproportionately low, the author can seek a retroactive adjustment to the contract.

If the "best-seller" rule is extended to workers' data, then this could have potentially far-reaching implications for industries where corporate automation projects are underway or planned. Those implications could further extend to workers and companies based outside the EU and therefore have an extraterritorial effect.[319] However, because the rules have seemingly been drafted to address another problem, there are a host of questions that remain unanswered. First, it is not clear how the economic value of workers' work products, which are protected by copyright, should change when transformed into data, and when aggregated within datasets. The value of data arises not only from quality, but also from volume and context, which are factors not easily addressed by traditional copyright

313. *See supra* Section VII.B.

314. *Id.* (referring to prop. 2021/22:278, *supra* note 40, at 159–60).

315. DSM Directive, *supra* note 10, at recital 81, art. 23(1). The optional nature of Article 18 of the DSM Directive follows from that it is not expressly referred to in Article 23 as being mandatory.

316. *Id.*

317. *Id.* at art. 20(1).

318. *Id.* at recital 73.

319. *See supra* Section VII.D. However, any right to redress that exists in these transnational circumstances will be limited to the EU-specific copyright portfolio and EU-specific use of the data, which makes valuation adjustments significantly more complex. This is discussed in more detail in Section IX.D.12, *infra*.

valuation methods. Second, it remains to be determined what threshold of economic disproportion should trigger the right to contract adjustment. The rule was never intended to apply universally, but only "in the event that the economic value of the rights turns out to be significantly higher than initially estimated."[320] Whether and when this threshold is met in the context of AI training data is an open question. It is also not clear how the rule should apply in the cases where the resulting AI system is used to replace the worker, but if traditional economic theories apply, it would suggest that the company may have extracted significantly greater economic value from the copyrighted works.

Contract adjustment mechanisms are, as discussed, nothing new in continental European copyright law, and have existed in several Member States long before the DSM Directive. Following Brexit, the UK chose not to implement the DSM Directive. There have subsequently been parliamentary discussions on whether to introduce a contract adjustment mechanism in the UK, similar to that of the DSM Directive. The Digital, Culture, Media and Sport Select Committee made several recommendations to the House of Commons in its final report in 2022, including a contract adjustment mechanism, arguing that this "would give creators greater leverage when negotiation [sic] contracts with music companies." [321] The UK government's response was that the "impact of these proposed new rights are uncertain and warrant further analysis." [322] The UK government has since commissioned research on these issues, which led to the UK IPO publishing a report in February 2023.[323] So far no new substantive proposals have been made in the UK, but a similar bill in the context of musicians and certain performers did recently pass. The Copyright (Rights and Remuneration of Musicians, Etc.) Bill amends the Copyright, Designs and Patents Act 1988 to introduce a right to equitable remuneration where the making available right is transferred concerning sound recordings, a transparency obligation, and a right to contract adjustment.[324] These rights would, however, not apply in the present context of workers and their work products being used as a data source.

Contract adjustment mechanisms are rare outside Europe. The United States has no similar contract adjustment mechanism. Moreover, because the work-for-hire doctrine in U.S. copyright law assigns copyright to the employer by default in employee–employer relationships, it seems even less likely that such a rule could come into play. That said, U.S. copyright law does contain a recapture right that shares some conceptual similarity. For works published after 1978, authors can terminate copyright assignments or licenses after thirty-five years, allowing them

320. DSM Directive, *supra* note 10, at recital 78.

321. DIGITAL, CULTURE, MEDIA & SPORT COMM., ECONOMICS OF MUSIC STREAMING: SECOND REPORT OF SESSION 2021–22, HC 50, at 67 (UK).

322. DIGITAL, CULTURE, MEDIA & SPORT COMM., ECONOMICS OF MUSIC STREAMING: GOVERNMENT AND COMPETITION AND MARKETS AUTHORITY RESPONSES TO COMMITTEE'S SECOND REPORT: SECOND SPECIAL REPORT OF SESSION 2021–22, HC 719, at 2 (UK).

323. U.K. IPO, *supra* note 178.

324. Copyright (Rights and Remuneration of Musicians, Etc.) Bill 2021–22, HC Bill [19] (UK).

to renegotiate terms or reassign rights.[325] The fact that this right only materializes after thirty-five years means that it will virtually never come into play for most workers, and even if it does, workers who have been made redundant from AI a long time ago will have little benefit from the rule. None of this is surprising as the right to recapture was mainly intended for other authors, such as musicians, artists, or performers, whose works are used and contracted for a long time.

D. A CONTRACT ADJUSTMENT ROADMAP FOR WORKERS' DATA

There is an urgent need to critically assess whether a contract adjustment mechanism is suitable for addressing the growing gap between the economic value workers generate through their data and the compensation they receive. The ability to adjust exploitation contracts could offer workers new legal recourse in the EU, where their work products are used to develop AI systems or drive automation. This would be especially significant for workers made redundant as a result of such technologies. However, no empirical studies have yet examined the potential economic impacts of applying contract adjustment mechanisms within employment or contractor settings. Nor has there been substantive discussion on how these mechanisms might apply to copyrighted works repurposed as data inputs for AI. These are unexplored issues of major concern, and the need to explore them is increasingly urgent as workers' data is expected to be used to a considerable extent in the coming years, potentially leading to the displacement of millions of jobs. It goes beyond the scope of this Article to analyze these issues in detail, yet the call for such analysis and discussion could not be greater. However, because there may already be an existing cause of action for workers whose data is used in the EU for AI or automation purposes, it is still meaningful to discuss what will be the most important questions to address, by way of an intellectual roadmap, in these contract adjustment cases. Twelve such important questions and various common scenarios are covered below.

1. Does the text and data mining exception exempt organizations from the contract adjustment mechanism?

The text and data mining exception in Article 4(1) of the DSM Directive requires Member States to introduce into national law exceptions and limitations for the reproduction and extraction of "lawfully accessible" works and other subject matter for the purposes of text and data mining.[326] Consequently, in the absence of an opt-

325. 17 U.S.C. § 203(a)(3).
326. DSM Directive, *supra* note 10, at art. 4(1).

out by the rightsholder under Article 4(3), the use of copyrighted works for text and data mining does not constitute copyright infringement in the EU.[327] The text and data mining exception raises the question whether companies can also rely on that exception in the case of workers' data. There is nothing suggesting that they may not. Companies will have lawful access to the works produced by their workers, and will frequently be the owners of any copyright attached to such works.

That companies may carry out text and data mining activities with respect to workers' data without infringing does not mean, however, that workers will not be entitled to claim further remuneration for such exploitation. The "best-seller" rule in Article 20 of the DSM Directive to adjust contracts assumes that the exploiting party has an existing, lawful right to exploit the works, whether by license or assignment.[328] The fact that there may *also* be a lawful right to exploit the works on the basis of statutory exceptions and limitations does not change this. A contract has still been entered into. In fact, it is the employment or contractor agreement which is what makes the company's right to exploit the workers' work products "lawful," unless copyright ownership is automatically assigned by statute in national law. If it were not for the employer–employee or contractee–contractor relationship, the original authors of the works would be in the same position as other rightsholders and have the possibility to reserve their rights by opting-out of text and data mining.

In short, the answer to the question is no: the text and data mining exception under EU copyright law does not exempt organizations from the contract adjustment mechanism. If workers suffer a substantial loss in the economic value derived from their work products, they may still be entitled to seek fair, additional remuneration under the DSM Directive.

2. Is an infringing act carried out in the EU in respect of workers' data?

The right to adjust contracts previously entered into for additional, fair remuneration is triggered only with respect to workers' EU-specific copyright portfolio and only insofar as a hypothetically infringing act that can be localized to the EU. This is a hypothetical question as the employer or contractee will have the right to use the workers' work products as a matter of contract or statute. Yet it is still important to answer initially, because what is fair remuneration for the EU-specific copyright portfolio can only be called into question by acts occurring in the EU. If no such acts occur in the EU, the EU-specific copyright portfolio is not implicated, and no entitlement to fair remuneration under EU copyright law arises.

Workers' data, like other data, will typically be reproduced as part of the training dataset used to develop an AI system or model. That act of reproduction will generally occur, at least, at the location of the server storing the data.[329] More

327. *Id.* at art. 4(3).
328. *Id.* at art. 20.
329. *See* Rättzén, *supra* note 9, at 232.

complex scenarios arise when the individual downloading or uploading the data is located in one jurisdiction while the server is located in another. In such cases, where multiple connecting factors exist, several states may each assert jurisdiction over the activity, consistent with the principle that states may apply their laws to conduct that takes place, wholly or partly, within their territory.[330] Recognizing this, the Canadian Supreme Court has famously held that, "[i]n terms of the Internet, relevant connecting factors would include the situs of the content provider, the host server, the intermediaries, and the end user . . . that Canada could exercise copyright jurisdiction in respect [of] both of transmissions originating here, and transmissions originating abroad but received here, is not only consistent with our general law but with both national and international copyright practice."[331] Similar views have also been favored by courts in the United States,[332] UK,[333] and Europe,[334] which have generally attached legal significance to all relevant territorial connecting factors. Accordingly, if neither the server nor the uploading/downloading individual is located in the EU, there can be no act of reproduction localizable to the EU—and thus, no basis for triggering the contract adjustment mechanism under Article 20 of the DSM Directive.

One possible exception arises where the person instructing or directing the training activities is based in the EU. This could become relevant where a company organizes their activities such that the infringing act of reproduction takes place outside the EU, but has operations in the EU which are participating in the training activities. The scenario can be compared with agency or management situations, where a principal instructs its own operations or an agent to commit infringing acts abroad. Jurisdictions differ on the treatment of such cases. In *Subafilms*, the Ninth Circuit held that authorization within the United States of infringing acts abroad did not trigger U.S. copyright liability.[335] Consistent with earlier decisions, the Ninth

330. *Id.*

331. *See* Soc'y of Composers, Authors & Music Publishers of Canada v. Canadian Ass'n of Internet Providers, [2004] 2 S.C.R. 427, 430 (Can.).

332. *See* Nat'l Football League v. PrimeTime 24 Joint Venture, 211 F.3d 10, 12 (2d Cir. 2000) (holding that a public performance or display includes "each step in the process by which a protected work wends its way to its audience" (citing David v. Showtime/The Movie Channel, Inc., 697 F. Supp. 752, 759 (S.D.N.Y. 1988)).

333. *See* EMI Records Ltd. v. Brit. Sky Broad. Ltd. [2013] EWHC 379 (Ch), at ¶¶ 35–38 (finding that the act of communication to the public can be localized to the UK where the uploading party is located in the UK, notwithstanding that the server is located abroad).

334. *See* Case C-5/11, Titus Alexander Jochen Donner, 2012 E.C.R. I-00000 at ¶¶ 26–27 (holding that the distribution right in copyright is characterized by a series of acts going "at the very least" from the conclusion of a contract of sale to the performance thereof by delivery to a member of the public, and that acts giving rise to a distribution to the public may therefore take place in a number of member states); Case C-192/04, Lagardère Active Broadcast, 2005 E.C.R. I-7199 at ¶¶ 46, 53–55 (holding that neither EU law nor international law prevents Member States from localizing the act of broadcasting within their respective territories if transmitters located in those states are used, even if the broadcast has minimal or no impact on the rights holder's exclusive market in those states).

335. Subafilms, Ltd. v. MGM-Pathe Commc'ns Co., 24 F.3d 1088, 1099 (9th Cir. 1994).

Circuit stood fast that the U.S. Copyright Act does not apply to foreign activities.[336] Case law is, however, not conclusive. Years before *Subafilms*, the Second Circuit found in *Update Art* that the U.S. Copyright Act may reach foreign conduct where the type of infringement permits further reproduction abroad.[337] Courts in other countries such as Germany [338] and Sweden [339] have also accepted that mere authorization of foreign infringing acts can amount to infringement domestically.

The situation resembles that of contributory infringement, where the contributory act takes place domestically, but the primary infringing act takes place abroad. The prevailing view from national courts is that the contributory act in a cross-border situation shall follow the law applicable to the primary infringing act.[340] Meaning that, if the primary infringing act is not actionable abroad, then neither can the contributory act be infringing at home. Recognizing this, Mr. Justice Arnold noted in *British Broadcasting* that "[c]onsistently with the territorial nature of UK copyright, any act constituting a primary infringement of copyright must take place within the UK."[341]

---

336. In particular, the Ninth Circuit explained that "we think it inappropriate for the courts to act in a manner that might disrupt Congress's efforts to secure a more stable international intellectual property regime unless Congress otherwise clearly has expressed its intent. The application of American copyright law to acts of infringement that occur entirely overseas clearly could have this effect. Extraterritorial application of American law would be contrary to the spirit of the Berne Convention, and might offend other member nations by effectively displacing their law in circumstances in which previously it was assumed to govern." *Id.* at 1097.

337. Update Art, Inc. v. Modiin Publ'g, Ltd., 843 F.2d 67, 73 (2d Cir. 1988). Several U.S. district courts have also declined to follow *Subafilms*. *See, e.g.*, Expediters Int'l of Wash., Inc. v. Direct Line Cargo Mgmt. Serv., Inc., 995 F. Supp. 468, 477 (D.N.J. 1998) (holding that "mere authorization of infringing acts abroad constitutes direct infringement and is actionable under United States Copyright Law"); Curb v. MCA Records, Inc., 898 F. Supp. 586, 594 (M.D. Tenn. 1995) ("*Subafilms* relies upon a peculiar interpretation of the scope and nature of the authorization right in 17 U.S.C. § 106. This interpretation . . . appears contrary . . . to well-reasoned precedent, statutory text, and legislative history."); Nat'l Football League v. Primetime 24 Joint Venture, No. 98 Civ. 3778 (LMM), 1999 U.S. Dist. LEXIS 3592, at *10 (S.D.N.Y. Mar. 24, 1999) (holding that, "where an individual commits an act of infringement in the United States that permits further reproduction outside the United States . . . a court may assert jurisdiction over those foreign acts and a plaintiff may recover damages for the infringing acts that took place extraterritorially.").

338. Bundesgerichtshof [BGH] [Federal Court of Justice] Mar. 29, 1960, GRUR 423, I ZR 109/58 (finding that an offer in Germany to sell infringing products in another country was considered patent infringement in Germany).

339. Svea hovrätt [HovR] [Svea Court of Appeals] Case T 1253/89, Dec. 12, 1990 (Swed.) (finding that when a Swedish company offered for sale a patented product to a foreign company, delivered from another foreign location, it constituted an infringing offer for sale in Sweden). However, in NJA 2005 at 180 (*Formsprutarna*) the Swedish Supreme Court found that the instruction from a Swedish company to manufacture infringing design products in another country did not constitute design infringement.

340. *See Subafilms*, 24 F.3d at 1092 (holding that "there could be no liability for contributory infringement unless the authorized or otherwise encouraged activity itself could amount to infringement"); Abkco Music & Records Inc. v. Music Collection Int'l Ltd. [1995] RPC 657 (UK) (holding that the act of authorization does not have to occur in the UK, provided that the primary act of infringement so authorized does); Bundesgerichtshof [BGH] [Federal Court of Justice], June 16, 1994 GRUR 798 I ZR 24/92 (Ger.) (holding that the act of authorization in Germany to conduct an auction of infringing products in another country did not amount to an infringement in Germany).

341. British Broad. Corp. & Anor v. Mech.-Copyright Prot. Soc'y Ltd & Ors [2018] EWHC 2931 (Ch).

The takeaway from all of this is that there is a risk that the "best-seller" right to contract adjustment in the DSM Directive, although it cannot be contracted out, could still be avoided by arranging the data training activities differently and locating them outside the EU. Even companies which are based in the EU could, in theory, exploit this potential loophole by sending workers' data to foreign servers and ensuring that only non-EU personnel are engaged in the training activities. Servers, and people, can of course be situated almost anywhere with little effort. Historically, however, European courts have shown a willingness to extend the reach of intellectual property laws where infringing conduct can be arbitrarily located somewhere else,[342] and particularly where the infringing party can still benefit from that foreign conduct in the forum.[343] This is no doubt the case for AI model training, and it remains to be seen whether courts would take a similar approach when it comes to reproductions of workers' data.

The main reason that extraterritoriality becomes a sensitive issue for using workers' data is because, more often than not, that data will be used to develop or deploy internal, not external, AI systems. In such cases, there will be no communication or distribution to the public, because there is no "public." However, to the extent that the AI system is communicated or distributed to third parties, and those parties are located in the EU, this gives rise to a second difficult issue: if training data is embedded within the model's parameters or weights, does the sale or licensing of the system equate to a sale of the training data? This is a highly complex technical question without clear answers.[344]

It is possible that some European courts might frame these contract adjustment claims not as primarily copyright-based, but as contractual in nature. Under this approach, the relevant nexus would be the contractual relationship, not the location of the reproduction. So long as courts have adjudicative jurisdiction over

342. *See* Case C-324/09, L'Oréal v. eBay, E.C.R I-6011 at ¶¶ 61–62; Case C-173/11, Football Dataco v. Sportradar, ECR I-0000 at ¶¶ 44–47. Similarly, in the patent context, the England and Wales High Court held in *Illumina, Inc v. Premaitha Health Plc* that there would be direct infringement in the UK for a method patent claim involving blood diagnostics, where the method was carried out abroad where there was no patent and where the results from the method were subsequently provided to customers in the UK. The High Court considered that the "substance" of the method was still performed inside the UK as blood samples were retrieved from UK customers, and "any other result would make it far too easy to avoid infringement of patents of this nature, given the ease of digital transmission and the ability to off-shore computer processing." Illumina, Inc. v. Premaitha Health Plc [2017] EWHC 2930 (Pat), at 507–08.

343. *L'Oréal*, E.C.R I-6011 at ¶¶ 62–63. A similar explanation was provided in Case C-173/11, Football Dataco v. Sportradar, ECR I-0000 at ¶¶ 44–47 (rejecting the argument that an act of re-utilization must be located exclusively to the territory of the member state where the web server is located from which the data in question is sent, which would impair the effectiveness of the protection afforded under that national law).

344. For a discussion generally about model memorization and encoding and copyright infringement, see Rättzén, *supra* note 9, at 220–27. In a recent German case, the Regional Court of Munich found that OpenAI infringed copyright by training its GPT language models with protected song lyrics without obtaining a license from GEMA. The court accepted that the lyrics had been "memorized" in the model's parameters and could be reproduced by simple prompts, constituting unauthorized reproduction and further acts communication to the public under German copyright law. *See* Case No. 42 O 14139/24, GEMA v OpenAI (Nov. 11, 2025) (Ger.).

the contract, then that may suffice from that point of view, regardless of copyright territoriality. In the EU, courts have special adjudicative jurisdiction in matters relating to a contract at the place of performance of the obligation in question.[345] The "place of performance of the obligation" in the Brussels Ia Regulation is understood to refer to, in the case of the sale of goods, the place where the goods were delivered or should have been delivered, or in the case of the provision of services, the place where the services were provided.[346]

In contrast, intellectual property infringement claims may be brought proceedings before the courts of the place where the harmful event occurred.[347] The concept of where the "harmful event" occurred has been interpreted to mean either where "the event giving rise to the damage" took place or where "the damage occurred."[348] It is unclear whether contract adjustment claims should be characterized as contractual or non-contractual under private international law. Although the claims concern the contractual relationship, the substantive right at issue comes from statute and serves to protect authors and ensure equitable remuneration for the exploitation of their works.

Where employment contracts are involved, additional jurisdictional rules may apply. Article 20(2) provides that, in matters relating to individual employment contracts, the employer shall be deemed domiciled in a Member State if it has a branch, agency, or other establishment there.[349] Disputes arising out of those operations may be brought in that Member State. Disputes over employment contracts may also be brought before the courts where the employee habitually carries out his work or in the courts for the last place where he did so.[350] The employment rules of special jurisdiction will become determinative if courts characterize the dispute as concerning the employment contract as such. This will turn on the specific facts, including the nature of the relationship and the terms of the contract.

3. Are all employment and contractor agreements exploitation contracts?

The right to contract adjustment only applies to contracts "for the exploitation of their rights."[351] What is meant by exploitation contracts is not further defined in

345. Regulation (EU) No 1215/2012 of the European Parliament and of the Council on Jurisdiction and the Recognition and Enforcement of Judgments in Civil and Commercial Matters, 2012 O.J. (L 351) 1 [hereinafter, the "Brussels Ia Regulation"], at art. 7(1)(a).

346. *Id*. at art. 7(1)(b).

347. *Id*. at art. 7(2).

348. Case C-68/93, Shevill v. Presse Alliance, 1995 E.C.R. I-415, ¶ 20.

349. Brussels Ia Regulation, *supra* note 345, at art. 20(2).

350. *Id.* at art. 21(1)(b)(i).

351. DSM Directive, *supra* note 10, at art. 20(1).

the DSM Directive.[352] As discussed above,[353] it was considered during the negotiations of the DSM Directive whether the rules should exclude employment contracts, which Sweden proposed. This was ultimately not accepted by other Member States as it could make it easier for the parties to avoid the rules by designating the exploitation under a particular contract label. Thus, the agreed premise was that all forms of exploitation contracts should be covered, including employment and contractor agreements.

Not all employment or contractor agreements are alike. In some cases, the license or assignment of copyright forms an essential part of the bargain. The employer or contractee has entered into the contract for the main purpose of obtaining the author's or performer's rights. In other cases, the license or assignment may be incidental or of limited value: the contract is primarily for the worker to perform ordinary duties, which sometimes will involve producing work products. The resulting copyright is transferred as a consequence, and although it possesses some value it is not the main purpose for entering into the employment contract.

There is an argument to be made that the statutory wording, contracts "for" the exploitation of works, suggests that the exploitation should form an essential purpose for the contract for the "best-seller" rule to apply. Yet, the fact that Member States choose not to define the scope any further on account of preventing the circumvention of the rules suggests that a broader interpretation should be favored. Limiting the rule to cases where exploitation is the main purpose for entering into the agreement would make it easy to sidestep. If that was the case, then a party in need of exploiting works could have the license or assignment clause included in a larger contractual context, thereby diluting its main or essential purpose. It is more likely that courts will come to interpret the rules broadly as encompassing any contracts involving any form of exploitation and to any extent. The relative extent and significance of the exploitation of the works is more properly analyzed in the substantive test for contract adjustment. Where the exploitation of the works is of limited value to the bargain between the parties, then courts will likely be more reluctant to find significant undervaluation that justifies adjustment.

4. When will workers' data yield a significantly higher economic value than their underlying work products?

The basic premise is that all authors and performers, even those in a work setting, should receive fair remuneration for the exploitation of their works.[354] This

352. Although not relevant for the present context, recital 78 refers to "contracts for the exploitation of rights harmonised at Union level," which may suggest that some contracts which concern exploitation *not* harmonized should be excluded. *Id.* at recital 78. *See also* BENTLY ET AL., *supra* note 164, at 48.

353. *See supra* Section VII.B.

354. DSM Directive, *supra* note 10, at art. 18(1).

translates into assessing what is the actual or potential economic value of the rights which are exploited, taking into account the author's or performer's contribution to the overall work or other subject matter and all other circumstances of the case, such as market practices or the actual exploitation of the work.[355] The right to have previously agreed contracts retroactively adjusted is a drastic and extraordinary legal measure. It should be the exception rather than the norm.

The right to have employment or independent contractor agreements adjusted will turn on whether workers' data yields a "significantly higher economic value" than their underlying work products.[356] If that is the case, and if workers have been "disproportionately" undercompensated when compared with the compensation provided and initially agreed upon, then workers will have the right to adjust their contracts retroactively. Not every discrepancy between what compensation has already been paid, and what greater economic value has not been paid for, will trigger this right. What is needed is a clear and significant discrepancy.[357] This is a high threshold to meet in most cases, and will require economic evidence and legal interpretation, a task for European courts to refine in the years to come.

Several scenarios may plausibly give rise to such a significant discrepancy between the compensation received by workers and the economic value realized from their data, which are discussed below. These scenarios are not mutually exclusive and can, and often will, overlap in practice.

Scenario 1: The parties have not intended or foreseen at the time of contracting the greater economic value which can be realized from workers' data.

There is no requirement that the full, realized economic value must have been unforeseen by the parties. However, the fact that neither party, especially the worker, anticipated the potential for such value may be a legally relevant factor.[358] A significant divergence between the parties' expectations and the actual outcome may signal that contract adjustment is warranted. This situation may arise, for example, where work products were never intended or discussed as potential data sources, or where their subsequent use as data yielded far greater economic value than originally contemplated. Recital 78 of the DSM Directive confirms that the parties' intentions and their anticipations are relevant factors to consider when assessing whether an adjustment is justified.[359]

355. *Id.* at recital 73.

356. *Id*. at recital 78.

357. *Id.* at recitals 73, 78. *See also* prop. 2021/22:278, *supra* note 40, at 176–77, 229–30 (stating that there must be a "clear discrepancy" between the contracted remuneration and the subsequent economic value for exploiting the works, which turns out to be "significant").

358. Prop. 2021/22:278, *supra* note 40, at 228–29.

359. Specifically, Recital 78 states that "[c]ertain contracts for the exploitation of rights harmonized at Union level are of long duration, offering few opportunities for authors and performers to renegotiate them with their contractual counterparts or their successors in title in the event that the economic value of the rights turns out to be significantly higher than initially *estimated*." DSM Directive, *supra* note 10, at recital 78 (emphasis added). When implementing the rules into national law, Sweden similarly

Scenario 2: Workers' compensation does not depend, or depends, on the extent of the use of their work products.

Another important factor to consider is whether the worker's compensation structure reflects the degree to which their work products are exploited. In some roles, work products may be of little importance, and mere byproducts to a broader set of duties which workers are being paid for. Simply put, the worker is primarily being compensated to do other things than produce work products, or the work products themselves have limited economic value. Income-based and impact-based valuation methods are less appropriate under these circumstances. If copyright law would intervene in such cases, then it could risk rewriting the original contract and the assumptions behind it. This is clearly not the intention behind the contract adjustment mechanism and would disregard the freedom of contract.[360] Rather, these mechanisms are meant to remunerate authors for their works so that the economic value is fully realized, as it should have been from the start in an equal and transparent bargaining situation. They should not be used as a legal vehicle to fundamentally change the underlying assumptions and terms of contract.

Scenario 3: The economic value of workers' data changes over time.

There will be many cases where work products will have been produced over longer periods of time, accumulating data. The parties may or may not have discussed or anticipated at the time of contracting that those products may be promising data sources for automation, particularly given how AI and machine learning technologies have only come to gain traction more recently. Importantly, the DSM Directive does not impose a fixed timeline or deadline for assessing the economic value of copyrighted works. Instead, Recital 73 confirms that remuneration should be "appropriate and proportionate to the actual or potential economic value of the licensed or transferred rights," which suggests that economic value realized much later in time, for other use cases, are equally relevant.[361] This interpretation is consistent with both income-based and impact-based valuation methods.[362] The right to adjust contracts is therefore both *ex ante* and *ex post*, in

emphasized the parties' intentions as relevant factors to consider. *See* prop. 2021/22:278, *supra* note 40, at 228–29.

360. Recital 78 appears to confirm this, stating that "[t]he assessment of the situation should take account of the specific circumstances of each case, including *the contribution of the author or performer*, as well as of *the specificities and remuneration practices in the different content sectors*, and whether the contract is based on a collective bargaining agreement." DSM Directive, *supra* note 10, at recital 78 (emphasis added).

361. *Id.* at recital 73.

362. Both income-based and impact-based valuation methods are concerned with the total economic value from a particular activity; it does not matter when (in time) the value arose so long as there is a causal relationship. *See supra* Section VIII.A (discussing income-based and impact-based valuation methods more generally).

the sense that all events occurring before and after contracting are relevant to consider when assessing the economic value. This is particularly relevant for workers' data accumulated over time, which may not be utilized until much later, possibly even after the worker has left the organization.

Scenario 4: Workers' data is used internally within the organization.

Although copyright law is primarily meant to incentivize the creation of creative content and prevent free-riding by third parties, there is no indication in the DSM Directive that works produced and used exclusively within organizations are excluded from its scope. Copyrighted works in employment and contractor settings are, as explained above, equally provided for under the rules.[363] This is because the contract adjustment mechanism is not necessarily designed to prevent market failure, but to ensure economic fairness and level the contracting playing field.[364] Consequently, the fact that the added economic value from using the works can only be traced to internal gains should not be a relevant factor as such.

When workers' data is used internally, such as to automate workflows, optimize processes, or support internal decision-making, the economic value may not be captured through direct revenue but through substantial cost savings and productivity gains. These benefits, though harder to quantify due to the absence of an external market price or benchmark, are nonetheless real and measurable. Both cost-based and impact-based valuation methods could be helpful to put a price on workers' data in these circumstances.[365] The most obvious example given in this Article is where the workers' data is used to replace workers themselves by automating their tasks.[366] There will be clear cost-savings for companies when doing so, and previously paid wages for those tasks may offer a reference point for valuing the data's contribution in such cases.

Scenario 5: Workers' data is used externally outside the organization.

Although it is uncommon for organizations to share or commercialize workers' data externally, given its strategic and often proprietary nature, such cases do arise. For example, companies may sell this data to third parties, license it for joint ventures, or use it in collaborative innovation projects. These scenarios typically involve a discernible price or exchange value, providing a clearer basis for economic valuation. Another common situation involves the use of workers' data in the development of new products or services that are marketed externally and generate independent revenue streams. In such cases, the data's contribution to external commercial success could potentially establish a case for a contract

363. *See supra* Sections VII.B and IX.C.

364. *See supra* Section VII.A (discussing the history and rationales behind the "best-seller" rule).

365. *See supra* Section VIII.B (discussing how cost-based and impact-based valuation methods apply to workers' data, including when such data is only used for internal purposes).

366. *See supra* Section VIII.C.

adjustment, especially where the worker's original contribution was not proportionately reflected in the remuneration received.

Scenario 6: Workers' data is used for producing new creative content.

Workers' data may not only support automation and analytical tasks but also serve as creative input in the generation of new artistic, literary, or design works. This is especially true in creative industries, where AI systems trained on a worker's prior outputs can be used to generate new content at scale. For example, AI trained on an illustrator's works may be capable of producing new images in a similar style. Even where the generated outputs are not direct copies, the underlying stylistic elements, such as shapes, color schemes, or composition, can be reused or recombined across diverse applications. This same logic applies to other creative fields, including journalism, music, marketing, branding, graphic design, software development, copywriting, and even professional services such as law and finance. In some cases, the AI-generated output may enhance existing workstreams, while in others it may substitute for or compete directly with the original human-created work.[367] There is added value in both cases, but arguably greater in the latter. Workers may be in a more favorable position for a contract adjustment case, when there have been no prior discussions about using their works as data sources, if the contract is unclear on the issue, or where the AI is used to displace workers and automate their previous job duties.

Scenario 7: Workers' data is used to come up with new inventions.

A more unique scenario arises where workers' data is used to develop AI systems or models that, in turn, lead to the creation of new inventions. This is especially relevant for workers in knowledge-intensive or specialist industry sectors. In the UK, an inventor is the "actual deviser" of the invention.[368] This means that it is the person who came up with the inventive concept who will become the inventor.[369] Other countries take similar approaches, with the United States for example defining inventorship by determining who "conceived the invention" or "contributed to the conception of the invention."[370] No court to this day has recognized AI systems themselves as inventors, if they are used as a tool to develop new inventions. To the contrary, the UK Supreme Court held in January 2024 that

367. *See* Rättzén, *supra* note 9, at 200–1 (discussing how AI-generated outputs could, in some cases, compete with original creators or reduce demand for their original works).

368. Patents Act 1977, c.37, § 7(3) (UK).

369. *See* Yeda Rsch. & Dev. Co. Ltd. v. Rhone-Poulenc Rorer Inter'l Holdings Inc. [2007] UKHL 43, ¶ 20.

370. *See Fiers* v. Revel, 984 F.2d 1164, 1168 (Fed. Cir. 1993); In re Hardee, 223 USPQ 1122, 1123 (Comm'r Pat. 1984).

patent law assumes that the inventor must be a natural person.[371] An AI is not a natural person, let alone a person. This of course leads to the question: who else can be the inventor, where AI has been used to generate the inventive concept or to materially contribute to it?

The U.S. Patent and Trademark Office published in February 2024 non-binding guidance on inventorship for AI-assisted inventions, which explained that:

> "A natural person who develops an essential building block from which the claimed invention is derived may be considered to have provided a significant contribution to the conception of the claimed invention even though the person was not present for or a participant in each activity that led to the conception of the claimed invention. In some situations, the natural person(s) who designs, builds, or trains an AI system in view of a specific problem to elicit a particular solution could be an inventor, where the designing, building, or training of the AI system is a significant contribution to the invention created with the AI system."[372]

It is not explained in the guidance what is meant by "essential building block," and what contributors are significant for "building" the AI system. It should not be categorically excluded that individuals who contribute data could be considered inventors. However, such cases are likely to be rare and would likely require that the data contribution played a decisive role in the invention's conception. In contrast, when large datasets are used and the AI independently identifies novel patterns, individual data contributors are less likely to meet the legal threshold for inventorship.

If workers who contribute their data to AI systems are, in fact, deemed inventors, then the laws for compensating employee-inventions will come to apply.[373] There is also a chance that they could claim further compensation in copyright law under the DSM Directive. The question will turn on what the added economic value of using their data is for the purpose of coming up with inventions, and to what extent workers have already received a fair portion of that value.[374] This analysis is necessarily fact-specific. Workers whose roles do not ordinarily include inventing, yet whose data has materially contributed to the invention, may be in a stronger position to argue for further compensation under a contract adjustment mechanism. This is especially true where their agreed remuneration did not contemplate such high-value, downstream uses of their copyrighted work, making it very similar to Scenario 1 above.

---

371. *See* Thaler v Comptroller-General of Patents, Designs and Trade Marks [2023] UKSC 49, ¶. 56. Similarly, in the United States, the Federal Circuit confirmed that an AI system could not be a valid inventor under U.S. patent law. *See* Thaler v. Vidal, 43 F.4th 1207, 1211 (Fed. Cir. 2022).

372. U.S. Patent & Trademark Off., Inventorship Guidance for AI-Assisted Inventions (Feb. 13, 2024), 89 Fed. Reg. 10043, 10049 (citing Dana-Farber Cancer Inst., Inc. v. Ono Pharm. Co., Ltd., 964 F.3d 1365, 1372–74 (Fed. Cir. 2020) (finding that "Drs. Freeman and Wood were found to be joint inventors even though they did not conceive of the claimed invention of using anti-PD-1 antibodies to treat tumors but instead discovered the expression of PD-L1 in human tumors and that PD-1/PD-LI interaction inhibits the immune response")).

373. *See supra* Section IX.B (discussing the law of employment invention compensation).

374. *See supra* Section VII.B.

Scenario 8: Workers' data is used to replace or displace workers.

Another scenario arises when workers' data is used to train AI systems that subsequently generate content or perform tasks that directly or indirectly compete with the original work products created by those same workers. This type of use can create substantial economic value for organizations. It allows for the computational scaling of creative or knowledge-based production processes, enabling the generation of more content, at a faster pace, and potentially of new types that would otherwise have been difficult or costly to produce manually. Depending on deployment, the AI system may perform functions similar to those of human workers, displacing them directly or gradually eroding the need for their roles.

The economic value of a worker's role is typically reflected in their compensation, which can serve as a benchmark when assessing the value an organization gains from automating that role. Cost-based valuation methods are especially relevant here, as they quantify the economic benefits derived from replacing or displacing labor.[375] Yet job loss alone does not entitle a worker to contract adjustment under Article 20 of the DSM Directive. Instead, what matters is whether the employer's gain is causally and substantially linked to the use of the worker's protected work, and whether it exceeds what was contractually anticipated and compensated for.

The worker's fair share of added economic value may vary with circumstances, and in some cases may be none. The fact that the worker has already been paid for their duties is not what is key in this regard, as has been explained. Rather, the question is whether there is a significant discrepancy between the agreed remuneration and the economic value later derived from the data. For example, a creative professional whose work is used to train an AI capable of producing similar outputs may be fairly compensated for the original content, but not necessarily for unexpected downstream uses of automation. If it cannot be shown that the agreed compensation contemplated such uses, a right to contract adjustment may arise.

5. How should the relative share of the economic value be apportioned to the use of work products as data?

The former analysis helped determine whether the use of workers' data in AI systems contributes to a significantly greater economic value than the value of the underlying work products. The next step, having concluded that there may in fact be such added value in some cases, is to ask how to calculate the workers' relative share of that value. In other words: what should be the additional price for using workers' work products as data sources?

375. *See supra* Section VII.B (discussing cost-based valuation methods as applied to workers' data).

The best measure of a good's economic value is often its market price.[376] Market prices are excellent proxies because they will have already factored in the supply and demand for the good.[377] They are the result of negotiations between different parties, and determine what others on the market are willing to pay. Actual market transactions serve as useful benchmarks for other sales in this regard.[378] When no external market exists, prices must be estimated based on related or similar goods, which is a classic problem in economics and one that courts have often addressed in intellectual property law.[379] When assessing damages for infringement, and in the absence of a clear licensing market price, courts have inferred value from comparable licenses.[380] If such licenses exist, entered into by other parties, their relevance must be assessed by comparing all relevant factors, including the licensed goods and their differences, the nature of the parties, the markets involved, the scope, territory, and duration of the license, and other terms that formed the basis of the negotiations.[381] Where no comparable licenses exist on the same market, courts have turned to related markets and constructed a hypothetical negotiation scenario between the parties.[382] This legal fiction imagines what a willing licensor and a willing licensee would have agreed upon under similar market conditions, but the outcome is inherently speculative.[383]

In most cases, there will be no direct market comparables for workers' data, as the data is unique to the organization. This makes its valuation more difficult. If a hypothetical license-fee approach is to be applied to workers' data, the question becomes what the organization would have been willing to pay for the data, if it were available for purchase on the market. The price of comparable datasets could assist in this regard, to the extent any such comparables exist. Where they do not, courts will need to construct the hypothetical market conditions themselves. This exercise involves distinguishing the received value of a good from the cost of obtaining or providing it.[384] This is easier said than done for data. What is the received value of particular data, especially when it forms only part of a larger dataset? And how should one assess the relative value of that data in relation to

376. *See* Daniel F. Spulber, *Access to Networks: Economic and Constitutional Connections*, 88 CORN. L. REV. 885, 899 (2003).

377. *Id.* at 895.

378. *Id.* at 900–1.

379. *See* Spulber, *supra* note 243, at 668.

380. *Id.* at 668–9.

381. Georgia-Pac. Corp. v. U.S. Plywood Corp., 318 F. Supp. 1116, 1119–20 (S.D.N.Y. 1970), modified sub nom. Georgia-Pac. Corp. v. U.S. Plywood-Champion Papers, Inc., 446 F.2d 295 (2d Cir. 1971)); Unisplay, S.A. v. Am. Elec. Sign Co., Inc., 69 F.3d 512, 517 n.7 (Fed. Cir. 1995); Smith Kline & French Laboratories Ltds (Cimetidine) Patents [1990] R.P.C. 203 (UK) (holding that "object of the comparability exercise, in this as in any other branch of the law, is to find the closest possible parallel. If there is an exact parallel, there is no point in looking any further. If there are slight differences, an allowance may be made."); Unwired Planet Int'l Ltd. v. Huawei Techs. Co., Ltd. [2017] EWHC 2988 (Pat), at ¶ 170 (UK).

382. *See* Spulber, *supra* note 243, at 622.

383. *See* John C. Jarosz & Michael J. Chapman, *The Hypothetical Negotiation and Reasonable Royalty Damages: The Tail Wagging the Dog*, 16 STAN. TECH. L. REV. 769, 783–84 (2013).

384. *See* Spulber, *supra* note 376, at 902.

the performance of an AI model? In practice, answering these questions would require a counterfactual baseline, comparing the performance of the model with and without the data.

The difficulty of assigning a price to data on the basis of hypothetical market values suggests the need to consider alternative valuation methods, such as cost-based, income-based or impact-based approaches.[385] Cost-based methods are attractive because they are quantifiable. For workers' data, the attributed cost would at minimum include the worker's compensation, proportionally adjusted to reflect the time spent producing the relevant data rather than other tasks. Additional costs could include those of processing, storing and curating the data, as well as the technical personnel involved in developing the AI system. The challenge, however, is that cost-based methods may systematically undervalue workers' data, especially where it is unique, irreplaceable, or unavailable from other sources. In such cases, income-based or impact-based methods may provide a better measure, since they assess the context of use, the data's purpose, and the quantifiable outcomes it enables. Yet these approaches carry their own risk of overvaluation. For example, the data may be used in a highly profitable project, but its actual marginal contribution to profits might be negligible. Overestimation also arises where workers' data constitutes only a small portion of a much larger dataset, or where its role in the functioning of an AI system is incidental.

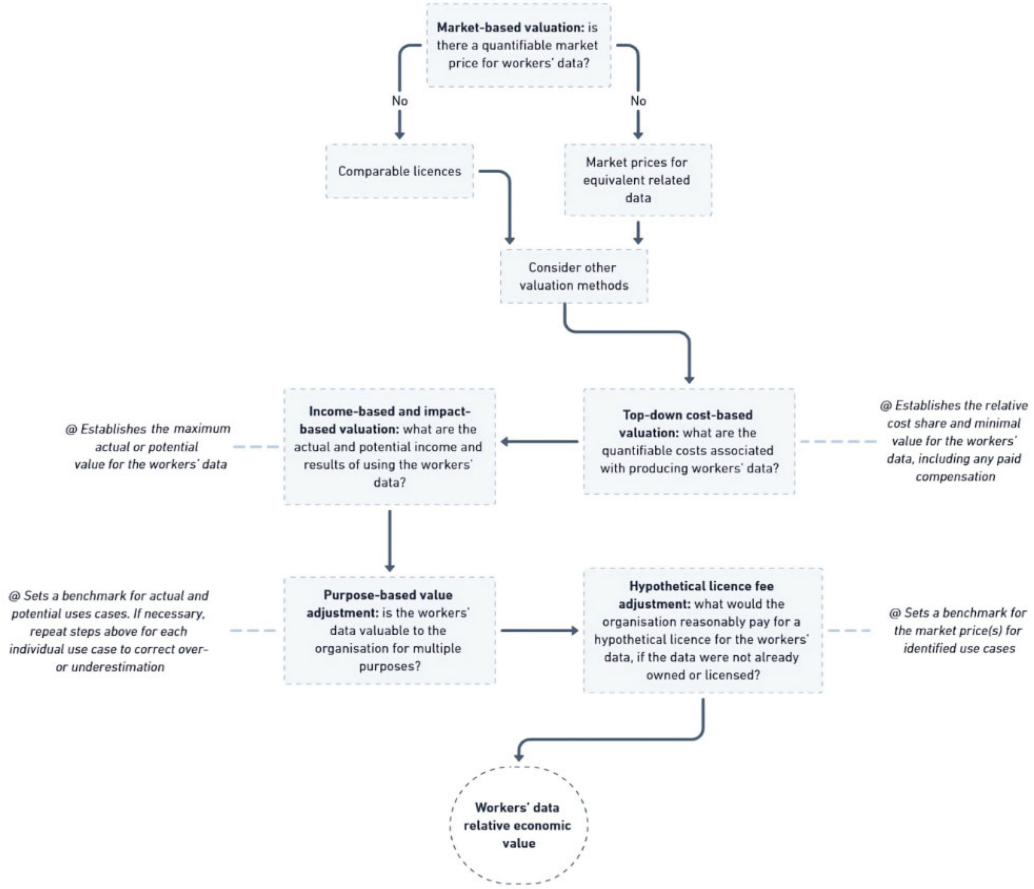
One way to navigate this dilemma is through a hybrid framework. A cost-based method can first be used to estimate the relative share of workers' data in the total costs of developing and deploying the AI system(s) in which it is used. This relative share provides a baseline, or *minimum* value as a "top-down" estimate.[386] Income-based and impact-based methods can then be used to establish an upper ceiling by estimating the *maximum* actual or potential value of the data in generating revenue or other gains. However, these methods must be applied carefully to avoid overestimating value. They may also underestimate the value of data when the data can serve multiple purposes that have not yet been realized. The value of a dataset is inherently contextual; it may be highly valuable in one application yet insignificant in another. The data should therefore be priced to account for all possible non-exclusive uses, in addition to the particular AI system which has already been deployed using the data. Once this analysis is complete, the final step would be to introduce a hypothetical licensing analysis: what would the organization be willing to pay to access the data, if it did not already own it? This hypothetical pricing scenario allows for correction of any inaccuracies from previous valuation steps and anchors the analysis in a more realistic market-facing perspective. The result is a

385. *See supra* Section VIII.B.

386. Top-down valuation methods are common in economics and have been discussed previously. *See supra* Section VIII.A. Although criticized, top-down methods have also been employed by courts and litigants in FRAND patent litigation. *See* Roya Ghafele, *(F)randonomics*, 20 COLO. TECH. L.J. 64, 79–87 (2022).

more balanced framework for estimating the relative economic value of workers' data, which is illustrated below.

The situation involving workers' data is unique in that the data is produced in the course of employment, and the worker has, more or less, already been compensated for generating it as a by-product of their work products. The bottom line should be that the fact that the worker has been a former worker should not put them in a better valuation position than where the data is sourced from a third party source. If anything, the fact that the worker has already received compensation, and may have assigned any copyright to the employer or contractor, suggests that workers should be paid *less* than what a third party licensor might receive. The proposed framework above can help courts understand the market



price of workers' data. But knowing that, courts must determine what should be the relative share of that market price, accounting for what has been previously

agreed between the contracting parties. Again, the relevant question is whether there is a "clear and significant" discrepancy between the economic value received by the employer and the remuneration received by the worker, one that renders the original compensation "disproportionately low."[387] The contract adjustment mechanism is *not* designed to fairly or evenly distribute economic value between authors and their contracting counterparts in all cases, but only where the economic value of their rights "turns out to be significantly higher than initially estimated."[388]

6. How should previously paid compensation to workers be factored in?

Authors should never be compensated twice for the same work. Companies may argue that their workers have already been fairly compensated for their data, having been paid for producing the underlying work products. While that may be true in some cases, it cannot automatically be assumed that compensation for the work products also covers the use of the data embedded within them. If that were always the case, then no contract could ever be adjusted. Indeed, a contract adjustment mechanism presupposes the existence of a contract and that the author has received some payment. The real question is whether the author was paid enough, taking into account *both* the economic value of the work and the terms agreed upon by the parties.

Recital 73 of the DSM Directive clarifies that a lump-sum payment can constitute proportionate remuneration in certain circumstances and that Member States should have the freedom to define specific cases for the application of lump sums, taking into account the specificities of each sector.[389] This implies that salaries, bonuses, and other forms of fixed compensation may count as fair remuneration, depending on the context. When implementing the DSM Directive, Sweden interpreted the rules to mean that compensation to authors in employment settings, which have been bargained in collective agreements, should be presumed to be fair.[390] However, the government made no statements about individual employment agreements that are not collectively negotiated and did not exclude the possibility of exceptions.[391] It also acknowledged that, even if an author is entitled to fair remuneration for the exploitation of their work under each contract, the "best-seller" contract adjustment rule allows consideration of factors beyond the specific contract in question.[392]

What will be key in these cases is whether the parties have discussed or intended that the work products may be used as data for automation purposes, or that there could be unexpected gains from using the work products. If so, the employer or

387. DSM Directive, *supra* note 10, at recitals 73, 78.
388. *Id.* at recital 78.
389. *Id.* at recital 73.
390. Prop. 2021/22:278, *supra* note 40, at 228.
391. *Id.*
392. *Id.* at 228–30.

contracting party has a strong argument that the agreed compensation already accounted for the data use.[393] A right to contract adjustment is also less likely to arise where the parties explicitly or implicitly agreed that the worker's pay would be determined solely by time worked or direct output, without regard to the actual or potential value of that work.[394] However, even where the worker has expressly or implicitly consented to the use of their work products as data, the contract adjustment rule does not exclude the possibility of further compensation. Because Article 20 of the DSM Directive is mandatory in nature, it cannot be waived by contract, meaning that the question still revolves around whether compensation already received is disproportionally low when compared to the realized economic value.[395] The parties may have underestimated the economic value at the of contracting, or the worker's bargaining position was weakened. The contract adjustment mechanism is also meant to step in to correct these imbalances.

7. Should workers or organizations reap the rewards for the higher economic value derived from their work products when transposed into data?

Although the legal principles underlying contract adjustment are well established, the rules were clearly drafted with a different scenario in mind. It is only recently that using data contained in copyrighted works for automation purposes has emerged as a new phenomenon, and it is not clear how the rules should apply in this context. The central issue is that while the individual work products may hold limited economic value on their own, when aggregated as data, they can generate substantially greater value. It is ultimately a policy question as to who should benefit from these additional rewards.

In this new context, companies may argue that one of the core responsibilities of workers in the age of AI is not just to produce traditional work products, but to generate data. From this perspective, the transformation of work into data is not a mere byproduct but an integral part of the job, and something workers are already being compensated for. Companies may also point to the significant investments and commercial risks involved in developing, maintaining, and deploying AI systems, arguing that these factors justify their entitlement to the bulk of the economic gains. The argument here is that these investments are not merely supportive but transformative; they turn raw data into a meaningful product that can generate significant economic returns in ways which the data alone cannot. Workers, by contrast, may argue that the economic value extracted from data ultimately derives from their creative labor. From that perspective, the data's usefulness, and its role in powering AI systems, stems directly from their intellectual contributions. From this standpoint, their right to fair remuneration should not be diminished simply because the medium has shifted from traditional work products

393. *Id.* at 229 (emphasizing the need to carefully consider what the parties have agreed upon before adjusting their contracts).

394. *Id.* at 229–30.

395. *Id.* at 229; DSM Directive, *supra* note 10, at art. 20.

to data, especially when those data-driven AI systems may ultimately replace their roles.

For now, EU copyright law appears to favor the workers' side of this debate. As discussed, Articles 18 to 20 of the DSM Directive are broadly framed and do not exclude any category of work, usage, or commercial context.[396] There is nothing in the Directive to suggest that the conversion of copyrighted works into data undermines the basic principle that creators should receive a fair share of the value generated by the exploitation of their works. If anything, the rise of data-driven business models introduces a growing risk that authors' contributions will be undervalued. Whether the emergence of these new uses calls for a revised legal framework, particularly in the context of employer-employee and contractor–contractee relationships, remains an open policy question.

8. How should the economic value of accumulated workers' data be assessed in relation to other data sources?

It is rare for AI systems to rely exclusively on data from a single source. More commonly, they are trained on data aggregated from multiple sources, which can include data from other workers, public databases or publicly available data, third party vendors or synthetic data. For that reason, workers' data is typically not used in isolation but as part of a broader dataset combining numerous sources. Data may have been sourced from the entire organization, including hundreds or thousands of other workers, along with other data sources accessible to the organization. Such data aggregation complicates the analysis of fair remuneration and the potential for contract adjustment. Once individual data contributions are combined, their value can become diluted, and direct attribution of economic benefit to a particular contribution becomes far more difficult.

From a company's perspective, once data contributions are aggregated into a unified dataset, the resulting asset is collective in nature. Hence, they may argue that the dataset should be assessed as a single whole rather than attempting to dissect and value individual contributions. The loss of granularity makes it impractical to link any specific worker's data to a measurable outcome or share of value creation. In contrast, workers may argue that individual contributions retain significance even when aggregated. Just as a copyrighted work retains its economic value when included in a database, a worker's contribution does not lose value simply because it becomes part of a larger dataset. The economic value of data in the context of AI is also closely linked to the quality, relevance and uniqueness of the data, not just its volume.

The reality likely lies somewhere in between these diverging positions and will depend on the context. Individual data contributions will become significantly harder to distinguish and put a price where they form part of larger models, such as large multimodal models and foundation models. However, companies

396. *See supra* Section VII.B.

increasingly use custom-built AI systems tailored for specific applications. These models are often fine-tuned on proprietary datasets generated by specific departments or teams.[397] For example, a company may develop internal AI agents trained exclusively on data from a particular department. In such cases, it becomes far more practical to estimate the economic value of individual data contributions.

Fairly distributing economic value to individual contributors whose contributions are used as a collective asset is a well-known problem in economics. The Shapley value can be used to resolve this problem, but will be computationally demanding, expensive and infeasible where the contributors are many.[398] Determining each contributor's marginal contribution means creating multiple counterfactual scenarios and excluding each contribution, one by one, until the relative impact of each contribution on model performance has been determined.[399] Variations of the Shapley value, such as Monte Carlo simulations, are arguably more appropriate for this purpose.[400] This will involve the sampling of a subset of possible coalitions of individual data contributors. Practically, this would mean estimating the average value of marginal contributions within the defined subset. This will not be completely accurate, but a more pragmatic choice.

Courts are not unfamiliar with economic estimations, and have frequently used them in FRAND patent litigation to estimate the essentiality ratio for patent portfolios.[401] A central issue in FRAND patent litigation is determining what standard-essential patents ("SEPs"), which have been declared as essential to a particular technology standard, are actually essential to that standard. This task is made more difficult when SEP portfolios are large, sometimes comprising thousands of patents from multiple parties, which would take an extremely long time to review individually. Once courts have determined what should be the aggregate license value for an SEP portfolio, they must allocate that value relatively among the individual SEPs.[402] Courts have gone about this in several ways. One such method is so-called patent counting or essentially counting, which establishes the total number of SEPs declared as essential, and then estimates the number of patents actually confirmed or deemed essential to the standard within that group.[403] This method crudely counts the number of patent contributions, without

397. *See supra* Section II (discussing common company use cases for workers' data).

398. *See supra* Section VIII.B (discussing the Shapley value and how it can measure the value of individual data contributions).

399. *Id.*

400. *Id.*

401. *See infra* notes 403–406.

402. *See* SEPs Expert Group, *Contribution to the Debate on SEPs* (Jan. 2021), at 108, https://www.pagewhite.com/images/content/SEPs_Expert_Group_Contribution_to_the_Debate_on_SEPs.pdf [https://web.archive.org/web/20250927011229/https://www.pagewhite.com/images/content/SEPs_Expert_Group_Contribution_to_the_Debate_on_SEPs.pdf].

403. *See* Unwired Planet Int'l Ltd. v. Huawei Techs. Co. Ltd. & Anor, [2017] EWHC (Pat) 711 (UK). In a report for the European Commission, it has also been proposed that analyzing random samples of declared SEPs would be an alternative for assessing whether the SEPs are truly essential. *See* Pierre Régibeau et al., *Transparency, Predictability, and Efficiency of SSO-Based Standardization and SEP*

assessing their actual importance and relative value, and has rightly been criticized for that reason.[404] Similarly, patent-weighted counting assumes that each patent in a given portfolio by a SEP holder has the same value. For example, if one SEP holder has charged a 10% royalty fee for a portfolio consisting of a hundred patents, the royalty fee is 0.1% per patent. The justification for doing so is purely practical, as it would be an unrealistic exercise to evaluate the individual value and relative significance of hundreds or thousands of patents.[405] A related method which has been proposed to determine the value of a patent portfolio is to look at the number of patents that have been accepted into a technical standard. For example, if more than 50% of all SEPs within a standard come from one patent holder, then this may suggest that this patent holder owns a large SEP market share and hence is entitled to a greater royalty share.[406]

None of these methods have been warmly accepted in patent law. They are very rough estimations only, without individually evaluating the significance of each patent. They could, however, be more promising in the context of workers' data. Unlike patents, which may differ significantly from each other in terms of relevance and validity, individual data contributions in datasets are likely to be more uniform and better categorized. Datasets are organized to account for different types of data sources, which is often a technical necessity for the model to work effectively. The value of the data will also be more tied to its relative volume. Data counting might be a promising method for estimating the value of individual data contributions forming part of a larger dataset, where data value is assumed to be relatively uniform. This approach assumes that each data contribution is equally useful across all use cases of the model. If the data can be separated into distinct subsets, then a more granular valuation may be possible. The following example illustrates how such a method might work in practice:

Facts:

A media company fine-tunes a third party LLM to generate automated news articles. The fine-tuning dataset comprises 20,000 articles written by 100 journalists, averaging 200 articles per journalist. All articles are categorized uniformly within the dataset. The AI system's effectiveness depends on the entire dataset rather than on

Licensing: A Report for the European Commission (June 2016), at 61; Justus Baron, *Counting Standard Contributions to Measure the Value of Patent Portfolios—A Tale of Apples and Oranges*, 44 TELECOMM. POL'Y. 13 (Apr. 2020). .

404. *Unwired Planet*, [2017] EWHC (Pat) 711 (UK); *see also* SEPs Expert Group, *supra* note 402, at 109.

405. *Unwired Planet*, [2017] EWHC (Pat) 711 at [183] (UK) (holding that "[s]hort of the disproportionate task of evaluating every single patent thoroughly in order to compare each one with all the others, one can only ever hope to analyze SEPs in broad categories and it is not meaningful to attempt to weigh the value of individual patents within these categories against one another").

406. However, this approach was rejected by the District Court for the Central District of California in *TCL Commc'n Tech. Holdings, Ltd. v. Telefonaktiebolaget LM Ericsson*, as it was based on the assumption that the SEP portfolio strength is directly proportional to its size in numbers and that each patent or contribution is treated equally. No. SACV 14-341 JVS(DFMx), 2017 WL 6611635, at *41 (C.D. Cal. Mar. 9, 2018).

any individual journalist's work.

Some journalists will have written more articles than others, and not all journalists are paid the same. Some articles from prominent journalists will have yielded more online viewings and therefore added greater revenue to the media company.

- Annual revenue from traditional (human-authored) articles: €2,000,000
- Annual revenue from the AI system using the dataset: €4,000,000

Uniform data contribution valuation:

If the value is distributed uniformly across all data points (articles), each article's contribution to the overall annual value as work products is €100 (€2,000,000 total revenue / 20,000 total articles). Therefore, each journalist's work product contribution value annually, by average, would then be €20,000 (€100 x 200 articles). The value of the work products as data sources would then be twice as much, €200 per article and €40,000 per journalist in total each year.

Individual data contribution valuation:

Let's adjust the calculations for the prominent journalists who could command higher revenue earnings for their articles as work products. If we assume that 10 prominent journalists would contribute double the annual revenue for their articles, then each article's contribution is €200 (€100 x 2). These 10 prominent journalists would, by average, contribute slightly fewer articles per year, approximately 175 per prominent journalist. Their total, combined annual value would therefore be €350,000 (€200 x 175 articles x 10 journalists), or €35,000 per prominent journalist.

The higher value of these articles as work products will not necessarily mean that a higher value should similarly be charged for them as data. For example, if the purpose of the AI system is to automate the writing of a journalist article, not necessarily in the style of a prominent author, if there is such a style, there is no reason why their data should be valued more than others. However, assuming that the fact that the articles are written by prominent journalists in a certain style is also valuable for the AI system, then the value as data would be twice as much as work products: €400 per article, €70,000 per journalist, and €700,000 in total AI-related annual revenue.

The remaining data value for the other non-prominent articles would then be: €4,000,000 - €700,000 = €3,300,000 total data value share. The average data value per non-prominent article would be €180 (€3,300,000 / 18,250 articles). Each non-prominent journalist would contribute with, by average, 203 articles per year (18,250 articles / 90 journalists). Each non-prominent journalist's data contribution would then be valued annually €36,540 (€180 x 203 articles).

While rough estimation methods such as data counting are unsatisfactory in patent law due to patent heterogeneity and complexity, they may be better suited for data valuation in copyright contexts. Unlike patents, copyrighted works are not subject to invalidation, and workers' data contributions are often well-structured and relatively uniform. Even where some contributions differ in quality or

prominence, it may be possible to distinguish their relative value, as illustrated above, by distinguishing their relative contributions and apportioning them a separate value. Altogether, this suggests that data counting is a relevant valuation method to consider for workers' data.

9. How should workers' redundancies be considered?

In principle, one's own loss is not relevant under economic valuation theories.[407] However, when a worker's loss results in a corresponding gain for the organization, that loss becomes economically significant. A parallel to this can be found in copyright law, where rightsholders are entitled to claim damages for lost profits.[408] That said, the contract adjustment mechanism in copyright law is not intended to fill potential policy gaps in employment law. It is not designed to compensate workers for job loss due to AI deployment. Rather, it aims to compensate workers for the additional economic value they bring to their former employers or contractual partners, both before and after being made redundant, that was not already reflected in their compensation.

As previously discussed, the economic value of workers' data, particularly in specialist and creative industries, is expected to be considerable when AI systems are used to automate tasks and replace human labor.[409] These systems can often perform the same tasks, and in some cases at a scale or speed that no human could match. One of AI's major advantages lies in its scalability, allowing it to perform work that would otherwise require many individuals. The value it generates will likely continue to materialize for several years after the original workers have been displaced. There is no built-in timeline or expiration date. Ultimately, this also means that companies that use or develop AI systems based on workers' data must carefully consider the implications of using such systems to replace their workforce.

10.Should different types of workers be remunerated differently for the

407. An exception to that premise is tort and restitution law, where the objective is to compensate for economic loss or to restore a previous economic state that corresponds to a social loss. *See* Francesco Parisi, *Liability for Pure Financial Loss: Revisiting the Economic Foundations of a Legal Doctrine*, *in* LIABILITY FOR PURE ECONOMIC LOSS IN EUROPE: FRONTIERS OF TORT LAW 75, 78 (Mauro Bussani & Vernon Palmer eds., 2003) (discussing the economic theories of tort law, including how only economic losses that constitute a social loss ought to be compensated).

408. *See* McRoberts Software, Inc. v. Media 100, Inc., 329 F.3d 557, 566 (7th Cir. 2003) ("[A]ctual damages are usually determined by the loss in the fair market value of the copyright, measured by the profits lost due to the infringement or by the value of the use of the copyrighted work to the infringer."). In the UK, claimants are able to claim either damages or accounts of profits; *see* Copyright, Designs and Patents Act 1988, § 96(2).

409. *See supra* Section VIII.C.

exploitation of their work products?

Copyright law rests on the principle that the law should not discriminate between different types of authors.[410] That all authors are treated equally does not mean that all works have equal value, however. Works can, and often do, differ significantly in their economic worth. The same logic applies to contract adjustment mechanisms. If an author has not been proportionally paid for their specific works, which may possess a greater economic value than others, then there is a claim to be made.

In the context of workers' data, the reality is that not all work products will carry the same economic value as data. Some workers will also have received higher pay than others because their contributions were deemed more valuable. There is nothing to suggest that Article 20 of the DSM Directive is meant to adjust those underlying assumptions. Rather, Recital 78 confirms that "[t]he assessment of the situation should take account of the specific circumstances of each case, including the contribution of the author or performer, as well as of the specificities and remuneration practices in the different content sectors."[411] In other words, the more valuable the work product is as data, the greater the potential economic value. However, and as noted above, the existence of high economic value in the use of work products as data does not automatically entitle the worker to additional compensation.[412] In each case, it is necessary to examine the specific terms of the employment or contractor agreement to determine whether that value has already been sufficiently compensated, and for what exactly. In some roles, the core duties may not involve producing economically valuable work products, or the work products may hold limited independent economic value. If copyright law were to adjust such contracts solely based on the value of the data generated, it would effectively rewrite the original agreement and its underlying assumptions. That cannot be the intended function of the DSM Directive's contract adjustment mechanism.[413]

If this understanding is correct, many workers would not have a claim to contract adjustment even if their data is used without further compensation, notwithstanding that the data can be of significant economic value for AI or automation purposes. This will particularly be the case for employees in administrative or support roles, where the primary duties are not to produce work products. Conversely, contract adjustment claims are more likely to arise in roles where the work product itself holds intrinsic value, such as those in knowledge-based, specialist, media or creative sectors. Which interpretation is correct in which scenario remains to be seen and will ultimately need to be decided by courts depending on the specific circumstances of each case.

410. *See* Christiana Sappa, *The Principle of Non-Discrimination*, *in* EU COPYRIGHT LAW 23–37 (Irini Stamatoudi & Paul Torremans eds., 2021).

411. DSM Directive, *supra* note 10 at recital 78.

412. *See supra* Section IX.D.4 (Scenario 2).

413. *Id.*

11.How should the worker be compensated for the future use of their data?

An extraordinary feature of the contract adjustment mechanism in Article 20 of the DSM Directive is that it accounts for *all* revenue gained from the licensed or assigned works, representing both "actual and potential" economic value.[414] This encompasses all past, present, and future income generated from the exploitation of the works. Whether this means that the author is compensated for future use will depend on the nature and terms of the original contract. As discussed in the preceding section, the purpose of the contract adjustment mechanism is to correct the distribution between the parties of the economic value, not to rewrite the underlying assumptions of the contract.

Where copyright has been assigned, future uses of the work product are necessarily included in the remuneration already paid, or in any additional compensation that may be awarded. Assignment is, by nature, a permanent transfer of rights. In contrast, where the copyright has only been licensed and royalties are paid periodically, any further remuneration will generally only cover past and ongoing exploitation, unless the license fee was structured as a lump sum. Only in such cases can remuneration for future use be contractually adjusted upfront. Since copyright is typically assigned to employers or contractors in employment or contracting relationships, assignment is the more relevant scenario for present purposes. The fact that contracts may be adjusted to account for not only past but also anticipated future revenue has significant implications. The task is to consider whether *all* of the economic value has been adequately accounted for in the original compensation. If not, and if the contract is not adjusted, that value will remain unaccounted for going forward. This is not a straightforward assessment and could lead to substantial additional liability for the exploiting party where the discrepancy in value is significant.

The difficulties of estimating future unrealized economic value will become even more pronounced for workers' data and AI. AI technology is relatively new, and predicting its development or market conditions a decade from now is highly uncertain. Additionally, companies often alter their business plans over time, or may face bankruptcy. It is not settled how such future events of uncertainty should be addressed in contract adjustment cases. It is also not clear what will happen if market projections are wrong. For example, if a court orders the adjustment of an assignment contract to better balance the economic value between the parties for exploiting the works, and if the assumptions made by the court regarding future exploitation turns out to be wrong, will the author become liable to repay to the other party what it should not have been paid? These are unresolved questions that future case law will need to address.

Consider the following hypothetical. Suppose a worker's contributions are used to train an AI system that is projected to generate €100,000 in annual revenue. Prior to being replaced by the AI, the worker received an annual salary of €50,000 as a

414. DSM Directive, *supra* note 10, at recital 73, art. 20.

lump sum for their work and copyright assignment. If the court finds this remuneration disproportionately low, and there was no mutual understanding at the time that the salary covered the use of their work as AI training data, what should be the appropriate time horizon for calculating future revenue owed to the worker? Should courts assume that the AI-related revenue will remain the same, or increase for the next five or ten years, or indefinitely at given market projections?

One possible compromise is for courts to only correct workers' fair remuneration, when in the form of lump sum payment such as salary, for past and present actions. Article 19 of the DSM Directive ensures that authors receive updated information about ongoing exploitation, including revenue data, at least annually.[415] Based on this, workers can claim additional remuneration periodically. Recital 75 clarifies that "[a]s long as exploitation is ongoing, contractual counterparts of authors and performers should provide information available to them on all modes of exploitation and on all relevant revenues worldwide with a regularity that is appropriate in the relevant sector, but at least annually."[416] Although more accurate, such a piecemeal approach is administratively burdensome and expensive. Practically, it would compel authors to repeatedly monitor exploitation, request data, and enforce their rights on a continuous basis to claim further remuneration for new exploitation, potentially indefinitely, so long as their works continue to be exploited. From a policy perspective, this is far from ideal, particularly since not all authors have the means, willingness, or legal resources to pursue such claims on a continuous basis.

12. How should the economic value of the EU-specific copyright portfolio and EU-specific acts be assessed when workers' data are used outside the EU?

The right to fair remuneration and the "best-seller" right to contract adjustment in the DSM Directive will only be triggered when there is EU-specific use concerning authors' EU-specific copyright portfolio. This is because copyright is a territorial right, as previously discussed.[417] The DSM Directive lays down rules for authors' bundle of copyright within the EU, which can only be infringed by acts carried out within the EU. Assuming a relevant act can be localized to the EU, the next question is how to put a price on the EU-specific copyright portfolio.[418] That analysis will depend on whether the relevant conduct is *also* carried outside the EU, or whether it is *only* carried out within the EU.

To estimate the value for the EU-specific copyright portfolio will be more difficult in the former situation, where the works, now in the form of data, are only partially or remotely used in the EU. There are two ways to approach that pricing dilemma. The first is to argue that the act of reproduction, even if only partially or remotely

415. *Id.* at art. 19.
416. *Id.* at recital 75.
417. *See supra* Section VII.D.
418. *See supra* Section IX.D.2 (discussing when an infringing act is deemed to be carried out within the EU).

occurring in the EU, serves as a jurisdictional "hook." Once such a hook is established, the rights to fair remuneration and contract adjustment should apply in full, even in respect of related acts conducted outside the EU. Although criticized, extraterritorial damages awards are not uncommon in copyright law, and have been awarded by American courts on numerous occasions.[419] Several European courts have taken similar views.[420] The alternative is to argue that the scope of the remedy should be in proportion to the amount or degree of infringing activity occurring in the relevant territory. This aligns more closely with the principle of copyright territoriality, in that remedies are only awarded with effect within the respective country of protection. Whether an act of reproduction occurs within the EU may also be arbitrary in the data lifecycle. Servers can be located anywhere, and therefore workers' data can be stored and processed anywhere. When data can be transmitted globally at any time, where exactly does the economic value of its reproduction lie?

A related but distinct view is that it is the economic value of the completed AI system, not the underlying data, that should determine value attribution. According to this view, the location of the AI system's use, whether inside or outside the organization, is what matters for apportioning value across regions or jurisdictions. This position is intuitively appealing because it aligns with the broader idea that the value of workers' data stems from its ultimate use in a functional AI system. It fits less neatly, however, with copyright territoriality if considered strictly. The relevant conduct from a copyright perspective for using workers' data will be the act of reproduction of the data, not its subsequent use within an AI system. Copyright laws do not grant exclusive rights to authors for the "use" of their works. Copyright laws are primarily concerned with copying, distribution, and performance rights. If we are to strictly conform to notions of territoriality and exclusive rights in this regard, we have to successfully argue that workers' data has been encoded or memorized into the AI system's parameters or weights, such that further acts of reproduction,

419. *See, e.g.*, Update Art, Inc. v. Modiin Publishing, Ltd., 843 F.2d 67, 68–70 (2d Cir. 1988) (holding that the predicate act of copyright infringement in the United States permitted the defendants' extraterritorial conduct, and therefore supported an award of damages for foreign lost profits); Sheldon v. Metro-Goldwyn Pictures Corp., 106 F.2d 45, 51 (2d Cir. 1939), *aff'd*, 309 U.S. 390 (1940) (awarding extraterritorial damages for foreign sales and distribution of an infringing motion picture produced in the United States); L.A. News Serv. v. Reuters Television Int'l, Ltd., 149 F.3d 987, 992 (9th Cir. 1998) (holding that the defendants were responsible for extraterritorial damages where predicate acts of infringement relating to broadcast videos took place in the United States); Tire Eng'g & Distrib., LLC v. Shandong Linglong Rubber Co., Ltd., 682 F.3d 292, 307 (4th Cir. 2012), *cert. denied*, 133 U.S. 846 (2013) ("[O]nce a plaintiff demonstrates a domestic violation of the Copyright Act, then, it may collect damages from foreign violations that are directly linked to the U.S. infringement.").

420. Bundesgerichtshof [BGH] [Federal Court of Justice] Dec. 19, 1975, ZR 110/74, 9 IIC 276, 276–81 (1977) (Ger.) ("The [defendant's] accounting on the distribution of the wall plates manufactured in the Federal Republic of Germany is to be without any territorial limitations since the [right holder's] royalty receipts payable by [its U.S. licensee] have been reduced through [the defendant's] distribution abroad."); Experience Hendrix LLC v. Times Newspapers Ltd. [2010] EWHC (Ch) 1986 [141] (UK) ("Losses might also be recoverable by infringement proceedings in respect of some overseas right brought in the jurisdiction where that overseas right is based.").

or communication or distribution to the public, will occur when the completed AI system is deployed within or outside the organization.[421] Alternatively, further acts of reproduction will occur where the generated output is identical or substantially similar to the underlying work products.

E. SHOULD WORKERS' PRIOR CONSENT BE REQUIRED OR AN OPT-OUT MECHANISM BE INTRODUCED FOR THE USE OF WORKERS' DATA?

The many difficulties with a contract adjustment mechanism and putting a price on workers' data raises the question whether another approach would be preferable. The starting point is that, reproducing copyrighted materials in the course of AI training activities, without the permission of the rightsholder, will infringe copyright unless there is an applicable exception or limitation.[422] There are significant differences in this regard between different jurisdictions in the absence of international harmonization on copyright exceptions and limitations.[423] In the EU, Member States must provide a text and data mining exception for reproductions and extractions of copyrighted works.[424] However, rightsholders may expressly reserve their rights, in which case their works cannot be used for text and data mining purposes.[425] This opt-out mechanism seeks to balance the interests of developers in accessing publicly available content with authors' interest in controlling the use of their works online.[426]

There are two ways to go about the same problem for workers' data:

1. **Prior and informed consent:** Workers would need to provide express consent before their work products could be reproduced or extracted for text and data mining, even if they have assigned their copyright to their employer or contracting entity.
2. **Statutory opt-out:** Companies may reproduce or extract workers' data for text and data mining by default, however workers would retain the right to exclude their work products, and hence data, through a positive opt-out mechanism, even after assigning their copyright.

Making the use of workers' data conditional upon the prior and informed consent of the worker, even if their rights for their copyrighted works, as work products, have been assigned to their employer or contractee is interesting for several reasons. It is, first of all, a clear and simple rule, and would bring clarity to

421. *See supra* Section IX.D.2 (referring to Rättzén, *supra* note 9, at 220–27, regarding a discussion about model memorization, encoding and copyright infringement).

422. *See* Rättzén, *supra* note 9, at 192–96 (discussing how AI training often involves carrying out acts of reproduction).

423. *Id.* at 196–220 (comparing the differences in data reproduction for training purposes under copyright exceptions and limitations in the EU, UK, U.S., Japan, Singapore, India, Australia, and Brazil).

424. DSM Directive, *supra* note 10, at art. 4(1).

425. *Id.* at art. 4(3).

426. *Id.* at recital 18.

workers and companies alike. It also requires companies to be transparent about if and how they plan to use workers' work products as data sources, before they do it. Such an *ex ante* approach would be strongly preferred over an *ex post* contract adjustment mechanism. Introducing an opt-out mechanism for workers' data would be advantageous for similar reasons, and ensure that workers are not completely at the mercy of their employers' or contractees' decisions regarding data use.

Requiring the prior and informed consent from workers for using their work products as data sources, even if copyright has been assigned, would come to mirror the legal treatment of personal data in several ways. Unless companies have a legitimate interest to process their employees' or contractors' personal data, or where it is necessary for the performance of a contract, it will often be necessary to obtain the consent of the data subject, which must be freely given.[427] Such consent is routinely provided in employment or contractor agreements. The risk that workers would refuse to provide their consent for using their workers' data is mostly theoretical. If a statutory consent regime is introduced, the greater risk is that such consent would become a mere formality if routinely expected as a condition of work, thereby reducing the rule to a transparency obligation with limited practical effect. While transparency is valuable, it does little to address the economic imbalance if workers' data is used to create automation systems that could ultimately replace their roles. Of course, in theory workers would be in a position to gauge if their employment or contractor agreement sufficiently compensates them for both their work products and their data. If that is not the case, then workers may either bargain for more compensation at the outset, or walk away. Yet workers are typically in a much weaker bargaining position than their respective employers or contractees. There is a significant risk that this could be taken advantage of at the cost of the weaker party.

The opt-out mechanism in Article 4(3) of the DSM Directive is currently limited to "rightsholders."[428] Workers who have assigned their rights to their employers or contractees will no longer be "rightsholders," meaning that the rule would need to be redrafted to extend to workers. Introducing a mandatory opt-out mechanism would put the control over workers' data back in the hands of workers. That is to say, even if copyright has been assigned or licensed in an employment or contractor setting, the worker would always have the statutory right to exclude the use of their work products as data sources. Because the worker would always have the last word. If an employer or contractee uses data despite a valid opt-out, the worker could claim fair remuneration through a contract adjustment. However, an opt-out regime introduces significant administrative and operational challenges. Workers would still need to take affirmative action. Passive workers who are not aware of their right to opt-out would be disadvantaged. Employers would also need to track opt-outs and trace corresponding data, which can be highly complex without robust

427. *See supra* Section IV.A (discussing data privacy restrictions for the processing of workers' personal data).

428. DSM Directive, *supra* note 10, at art. 4(3).

data labeling systems. If opt-outs were widely exercised, data scarcity could arise, potentially stifling corporate innovation and automation. Although the prohibitive effects of an opt-out mechanism could be a relatively rare occurrence, that is not to say it could not happen, and if it does, then it would be very serious for a company's competitiveness. Given the growing importance of data in a digitalized corporate environment, a blanket opt-out mechanism would be inappropriate. Clearly, companies need reliable access to the data within their own organizations.

A more balanced solution would be to introduce a compulsory licensing or compensation scheme. Unlike Article 20 DSM's case-specific contract adjustment mechanism, a compulsory license would standardize the right to further remuneration while allowing parties to negotiate specific terms. But standardize how? The truth is that most compulsory licensing statutes do not set out any fixed royalties or amounts, and leave it to the parties to negotiate what is fair compensation.[429] The scheme is usually limited to setting out the circumstances for when there is a statutory duty to provide a license, without setting out the details of such a license, which are best reserved for the parties to determine. There is much truth to that for also workers' data, which economic value will vary significantly depending on its content, use and context.

The existing "best-seller" right to adjust contracts can be highly accurate when determining what is a fair distribution of the economic value of workers' data, and is, in a way, a form of a compulsory licensing scheme. It requires companies exploiting authors' works to provide further remuneration, in addition to what has previously been agreed upon, where this is economically justified. Yet this is a highly fact-intense and complex analysis, as discussed above, which leaves open the question whether it should be standardized in some way. It goes beyond the scope of this Article to explore these questions any further, and it may well be that we will have to wait and see how workers' data negotiations pan out before further statutory intervention.

F. SHOULD A WORKERS' DATA EXCEPTION AND LIMITATION BE INTRODUCED?

Companies need reliable access to the data within their own organizations, particularly as corporate automation projects become both more common and more critical for maintaining competitiveness. The current contract adjustment mechanism in the DSM Directive, when applied to workers' data, does not restrict companies' right to use their workers' data when they are the lawful copyright owners.[430] However, this mechanism carries the risk of potentially having to pay workers additional remuneration, thereby raising the potential overhead costs of

429. *See* CMS, *The Compulsory Licensing E-Guide: An e-guide from the Global Intellectual Property Group* (Feb. 2021), https://cms.law/en/media/expert-guides/files-for-expert-guides/cms-compulsory-licensing-global-expert-guide-feb-2021 [https://web.archive.org/web/20251011015423/https://cms.law/en/media/expert-guides/files-for-expert-guides/cms-compulsory-licensing-global-expert-guide-feb-2021].

430. DSM Directive, *supra* note 10, at art. 20.

internal AI development and deployment. Valuing workers' data is also legally and economically complex. Nearly every contract adjustment case would involve a difficult and fact-intensive assessment of the economic value of the data, creating fertile ground for disputes and litigation. That in itself could be sufficient to impede or slow down corporate automation projects.

Given these challenges, it is worth considering whether a workers' data exception or limitation should be introduced. Article 4(1) of the DSM Directive already obliges Member States to implement exceptions and limitations for reproductions and extractions of works for text and data mining purposes.[431] This extends to workers' data, as previously discussed.[432] However, this does not mean that organizations are entirely free to use their workers' data, even when they hold the copyright for the underlying work products. The "best-seller" rule in Article 20 to adjust contracts previously entered into still applies to the contracts for the exploitation of their rights.[433] This currently puts companies in an awkward legal situation. They are not free to use the data within their own organization which is derived from their workers without being subject to the potential liability of having to pay further remuneration to those workers. Yet companies are free to use publicly available copyrighted works from third party rightsholders for text and data mining purposes. No compensation has to be paid to rightsholders in such circumstances. But of course, this statement is too simple, and fails to consider that third party rightsholders may have opted-out of text and data mining by reserving their rights under Article 4(3). If that is the case, then the choice companies have is really between "not using at all" third party data sources and "paying some" for workers' data.

The question is whether this is an acceptable compromise. One possible reform would be to redraft Article 4(1) so that it overrides Article 20, allowing companies to use their workers' data for text and data mining purposes without any obligation to provide further remuneration, regardless of the economic value of that exploitation. However, bluntly excluding text and data mining uses from Article 20 could have significant and unintended consequences, many of which are difficult to foresee at this stage. It could lead to a situation where workers' contributions are significantly undervalued, or where companies exploit their workers by a "hire and fire" approach. That is to say, companies employ workers to generate data for automating their tasks and then dismiss them once that data has served its purpose. Such job disposability could exacerbate what is already, in many cases, an unequal bargaining relationship to the detriment of workers. If a workers' data exception and limitation were to be introduced, it would need to strike a careful balance between facilitating corporate innovation and protecting workers' rights. One middle ground could be to allow companies to rely on a "legitimate interest" basis, similar to the GDPR, which would permit the use of workers' data for text and data

431. *Id.* at art. 4(1).
432. *See supra* Section IX.D.1.
433. *Id.*

mining without additional remuneration, provided certain transparency and fairness safeguards are met.[434] For example, such a "legitimate interest" rule could permit companies to transparently use workers' data for automation purposes which do not put workers' jobs at risk, but if disproportionately impacting the worker it could give rise to additional compensation.

Ultimately, only time will tell if the current system is skewed too much in favor of workers than companies who need reliable and effective access to workers' data. If that turns out to be the case, we will no doubt notice it. If companies with operations based in the EU struggle to find enough data for their AI or corporate automation projects, or pay too much for it, they will need to make their voices heard to policymakers. If those voices fall on deaf ears, then the EU may come to experience corporate migration in the thousands.

X. WHAT CAN WORKERS DO TO REALIZE THE VALUE OF THEIR DATA?

Workers today find themselves in a difficult position. The companies they work for increasingly rely on their work products as valuable data sources for developing and deploying AI systems. Not only do workers generally receive no additional compensation for this use of their data, but they may also face redundancy if their own work is used to automate the very tasks they perform. Millions of jobs across the world are expected to be lost in the coming years this way. The pressing question is: what can workers do to protect their interests and realize the value of their data?

As discussed throughout this Article, workers currently have few legal tools to rely on in most jurisdictions. Traditional contract and employment laws are not well equipped to address the challenges of a looming "job apocalypse." A notable exception exists in the EU, where the DSM Directive provides workers with the right to fair remuneration and the ability to seek contract adjustment in specific circumstances. However, this right is exceptional and reserved for cases where "the remuneration originally agreed turns out to be disproportionately low compared to all the subsequent relevant revenues derived from the exploitation of the works or performances."[435] In practice, the employer must have realized economic value "significantly higher" than what was initially paid or expected.[436] This means that what has been agreed and discussed between the parties becomes crucial. Although the right to adjust contract is indeed meant to do exactly that, adjust what has been agreed between the parties, what has been mutually intended will impact whether the remuneration is "disproportionately low." It is, therefore, both an objective and subjective analysis combined.

This framework has important implications for workers when negotiating employment or contractor agreements. Clarity will be king, and workers should

434. GDPR, *supra* note 48, at art. 6(1)(f).
435. DSM Directive, *supra* note 10, at art. 20.
436. *Id.* at recital 78.

carefully scrutinize any copyright assignment or license provisions. If the contract is silent or ambiguous on whether work products can be used as data sources, workers may wish to raise the issue and insist on that the contract explicitly defines the permissible uses of work products and the context in which they may be exploited. Workers who produce valuable data should also maintain evidence of what is, and is not, said during contract negotiations. While jurisdictions differ on the admissibility of pre-contractual discussions for contract interpretation,[437] preserving this evidence strengthens the worker's overall position. After entering into the contract, workers are entitled under the DSM Directive to receive "relevant and comprehensive information on the exploitation of their works [. . .] in particular as regards modes of exploitation, all revenues generated and remuneration due," at least annually.[438] The transparency obligation provides workers with the legal means for knowing how their works are exploited, which would include using their works for automation purposes. If and when an AI system has been developed using workers' data, workers are also entitled to continuously receive information about its revenue.

Collective action offers another option. Workers should consider leveraging unions or collective bargaining to address the use of work products as data sources. This would not be a new trend. For example, in May 2023, screenwriters in the United States went on strike, demanding that generative AI tools support, rather than replace, human writing.[439] The resulting agreement required that AI not be used to generate source material and affirmed that writers could choose whether to use AI as a tool. Importantly, that same agreement also stated that the Writers Guild of America West reserved the right to "assert that exploitation of writers' material to train AI is prohibited by MBA or other law."[440] The use of workers' data is an industry-wide issue and should, ideally, be resolved as such, rather than forcing workers to negotiate such difficult questions themselves on an individual basis. It is expected that there will be future class action lawsuits, if not union disputes, revolving around the use of workers' data, particularly where large groups of workers have been made redundant due to AI.

Finally, workers should recognize that their copyright arises automatically in nearly every jurisdiction, not only in the country where they reside or where their company is located. Even if they live or work outside the EU, the DSM Directive's

437. In the UK, for example, the general rule is that pre-contractual discussions are inadmissible as evidence to assist in determining the construction of a contract. *See* Prenn v. Simmonds, [1971] 1 W.L.R. 1381 (H.L) (UK).

438. DSM Directive, *supra* note 10, at art. 19(1).

439. *See* Alissa Wilkinson, *The Looming Threat of AI to Hollywood, and Why It Should Matter to You*, VOX (May 2, 2023), https://www.vox.com/culture/23700519/writers-strike-ai-2023-wga [wayback tk].

440. *See* Writers Guild of America West, *Summary of the 2023 WGA MBA* (Oct. 24, 2023), https://www.wga.org/contracts/contracts/mba/summary-of-the-2023-wga-mba [https://web.archive.org/web/20251011020747/https://www.wga.org/contracts/contracts/mba/summary-of-the-2023-wga-mba].

right to contract adjustment could still be relevant, as discussed in this Article.[441] For workers employed by multinational companies, or by organizations that conduct business in the EU, AI systems trained using their data could trigger a right to contract adjustment if that data is exploited in the EU. While the exact scope of these claims remains unsettled and may require judicial clarification, it is possible that only a modest degree of "use" of workers' data within the EU could be sufficient to establish a valid claim.

XI. WHAT CAN COMPANIES DO TO PROTECT THEIR INTERESTS?

Companies have a strong and legitimate interest in maintaining full access to the data generated within their organizations. Lacking such access could cripple business operations and corporate innovation, and increasingly so in a future driven by AI. The right to contract adjustment under the DSM Directive is expected to be triggered in more exceptional rather than routine cases, where workers' data forms a uniquely critical component of an AI system or model. Many routine automation projects will remain unaffected, particularly where the workers' data is generic, widely available, or less reliant on the intrinsic value of the content as a creative work. However, the risk of contract adjustments increases significantly when the data originates from knowledge workers, specialist employees, or individuals in creative and media industries. The risk is also heightened when the same work products are used to train AI systems that automate the very tasks those workers perform, especially if the employees are later made redundant. Similarly, concerns arise where workers' data is used to generate new inventions or to automate creative processes that produce commercially valuable content. In such scenarios, the economic value derived from the additional use of workers' outputs can potentially be substantial.

What has been agreed between the parties will similarly become critical in these cases. While the statutory right to contract adjustment cannot be waived, courts will consider what the parties intended and agreed upon when assessing whether the remuneration is "disproportionately low." Companies can, therefore, reduce risk by influencing relevant factors through careful drafting. Contracts may expressly state that employees' or contractors' work products may be used for text and data mining or similar purposes. Where a worker has been clearly informed of such use from the outset, courts are more likely to consider the original remuneration fair, even if the data subsequently generates significant economic value. That said, explicit contractual notice does not eliminate the risk of adjustment. To further protect their position, companies may also consider offering redundancy compensation packages to employees whose roles are automated using AI trained on their own work products. Such redundancy compensation could assist companies with arguing that workers who have been made redundant have been fairly remunerated for the use of their data.

441. *See supra* Sections IX.D.2 and IX.D.12.

Companies using or considering using workers' data as data sources should establish data governance policies that clearly define how the data is collected, stored, processed, and shared. Because the right to contract adjustment could come into play even for companies based outside the EU, it will equally be important to detail where workers' data are stored and used. Internal transparency regarding workers' data will be important not only for reducing the risk of contract adjustment claims, but also to comply with statutory transparency obligations. Article 19(1) of the DSM Directive imposes a proactive duty on companies to provide authors and performers with "relevant and comprehensive information" on the exploitation of their works and the revenues generated, at least on a regular basis.[442] This obligation applies regardless of whether the author formally requests the information, and there is no indication that works used solely for text and data mining are exempt.[443]

The legal situation is less clear when workers' data is used for coming up with new inventions. There is a lack of guidance and case law on whether the workers responsible for producing the relevant data could qualify as inventors in patent law, and companies are wise to account for that uncertainty when naming inventors and filing for patent applications.[444] If a worker would coincidentally be considered an inventor, due to the importance of their underlying data for coming up with the invention, then they will be entitled to claim inventorship and may be statutorily entitled to compensation in patent laws. There is also the possibility that they may be entitled to compensation on a contract adjustment basis in copyright law.

XII. LAST WORDS TO REGULATORS

The governance of workers' data has so far been overlooked in regulations across the world. Currently, the default position is that employers or contracting parties, either by statute or contract, hold the copyright in works created by individuals in the course of their duties. This long-standing rule has provided companies with the freedom and flexibility to use employees' work products as they see fit. In the era of corporate automation, this includes using those work products as data sources, which for many companies will be the most valuable data to use for AI and automation purposes.

The EU is a noteworthy exception. The DSM Directive introduced a right for authors to receive fair remuneration for the exploitation of their works, and a right to seek contract adjustments where that remuneration is disproportionately low. The legislative intent at the time was to address inequitable relationships and correct market failures where authors were undercompensated for the economic value of their works. It is, perhaps, a mere coincidence that the DSM Directive has been drafted in such a way that enables potential contract adjustment claims in

442. DSM Directive, *supra* note 10, at art. 19(1).
443. Prop. 2021/22:278, *supra* note 40, at 232.
444. *See supra* Section IX.D.4 (Scenario 7).

relation to workers' data. Although the legislative history of the Directive confirms that the intention has been for employment and contractor relationships to be covered by the rules, not a single word has been said about the use of copyrighted works as valuable data sources for AI and automation purposes. Nevertheless, the same economic logic applies. This Article has concluded that, if workers' data can generate significantly greater value for companies, especially when used to develop AI systems that automate the very tasks those workers performed, courts may find a basis for contract adjustment in certain cases. This will be particularly relevant in industries where the underlying work products themselves have high value, such as in knowledge-intensive, media, or creative sectors, and where workers are displaced as a result of the AI trained on their work. Though the existence of a "workers' data value gap" will always need to be assessed on a case-by-case basis.

We lack reliable evidence on how workers' data is currently being used. Although this is not surprising, as we are concerned with proprietary data, it makes it much more difficult to assess the scale of the issue or the urgency of a policy response. However, industry trends and news suggest that many companies have already begun, or will soon begin, replacing parts of their workforce with AI. If workers' data is a critical ingredient in these AI systems, we may come to see a new wave of litigation, particularly in the EU for contract adjustment claims. Whether this will strike the right balance between protecting workers and encouraging innovation remains to be seen.

There is, in principle, nothing extraordinary about rewarding workers for valuable contributions to their employers. Employee-invention statutes in patent law have compensated employees for their inventions to their employers for many years, even if the right to patent the inventions has been assigned to their employers and even if they have received compensation contractually for their work. What sets workers' data apart from workers' inventions, however, is that the data will often derive from work products produced in more routine tasks. There is a policy reason to incentivize new inventions, but there is no need to incentivize routine tasks which workers are already paid for. The rationale behind a right to contract adjustment is in many ways different. Instead, it serves as a protectionist instrument to correct market inequalities and even the contracting playing field, where the author often has a weaker bargaining position. That logic extends to many employment and contractor relationships as well, but arguably becomes more complex in the case of workers' data. What makes workers' data unique in many ways is that its value does not necessarily have to correlate with the value of the underlying work product. Depending on the context and application, its value may come from accumulation in the masses, often together with data from other workers. The economics behind workers' data is largely unexplored and this Article has identified a clear demand for further economic research. What is also not clear is if workers actually think that they should be paid more for their workers' data when adding greater value to the companies they work for, or if workers are mainly concerned with using their data to make their own jobs redundant. There is a need to conduct large-scale interviews to obtain this important policy information.

The right to contract adjustment is in many ways a blunt legal instrument. It is designed to address market inequalities *ex post*, relying on fact-intensive, complex and costly discovery and/or litigation. Article 21 of the DSM Directive accounts for that by requiring Member States to offer voluntary alternative dispute resolution mechanisms, but these have little practical impact if parties can simply refuse to participate.[445] A system that forces each dispute to be litigated individually is unsatisfactory for both workers and companies.

This Article has discussed other alternative legal mechanisms to tackle specifically the use of workers' data. One option would be to create a mandatory opt-out mechanism from text and data mining activities for workers' data, similar to what already exists in the DSM Directive for rightsholders who have not assigned their rights to someone else.[446] This would give workers control over their data and ensure they are not at the mercy of their employers' decisions. The downside of that is that companies could run into data scarcity issues, if all their workers decide to opt-out of using their work products as data. It also requires workers, who may not be aware of their rights, to take affirmative action by opting-out. Another option would be to treat workers' data similar to personal data and require companies to obtain prior and informed consent from the worker to use their work products as data sources for text and data mining purposes. This could however in practice turn out to favor companies at the cost of workers, if such consent provisions become boilerplate in most employment and contractor agreements. A hybrid approach of resolving that would be to condition the use of workers' data on certain "legitimate interest" factors. For example, companies could be allowed to use the data without workers' consent or having to pay additional remuneration, if they are transparent about such use and if it does not disproportionately disadvantage the worker.

Finally, regulators must confront the novel problem of workers' data being used to make workers themselves redundant. While technological displacement is not new, what is new is that employees' own work products, often protected by copyright, are now being used to build the very systems that replace them. If no legal recourse is provided for workers in these circumstances, there is a risk that some companies come to exploit this asymmetry by hiring workers to produce the data they need for automation purposes and then dismiss the same workers from their jobs. This creates a serious policy concern that warrants regulatory attention, whether through a right to contract adjustment or other legal means.

XIII. CONCLUSION

This Article may come as a wake-up call to companies that have initiated or completed corporate automation projects using workers' outputs as data sources. Conversely, it may offer some hope to workers who have lost, or may soon lose,

445. DSM Directive, *supra* note 10, at art. 21.

446. *See supra* Section IX.E.

their jobs to AI. This Article has examined what legal recourse workers may have against their employers or contractees when their work products are used for AI or automation purposes. While the outlook in most jurisdictions remains bleak, where employment, contract, and AI laws offer little protection, copyright law may provide an unexpected remedy to workers. In the EU, the right to contract adjustment, also known as the "best-seller" rule, under the DSM Directive could alter the legal analysis of using workers' data and potentially require companies to provide further compensation to workers.

This Article has concluded that a workers' data value gap could, in certain cases, arise where companies derive significantly greater economic value from using workers' work products as AI training data than what any agreed compensation reflects. Whether or not this triggers a contract adjustment claim under the "best-seller" rule will depend on a highly fact-specific analysis of the economic value generated versus the remuneration received and agreed upon. The right to contract adjustment is mostly a European legal mechanism, at least for now, and does not exist in the UK or the United States. European copyright law could however come to apply extraterritorially in cases where non-EU based companies use their workers' data for corporate automation projects deployed in the EU, where workers' data is stored or processed on servers located in the EU, or where the workers themselves are based in the EU.

In the coming years, the right to contract adjustment is expected to generate new litigation, both individual and collective, as millions of workers face AI-driven displacement. Whether this is a positive or negative development remains an open policy question, and it is uncertain whether copyright law provides the optimal framework to address the challenge. This Article urgently calls for economic and legal research to explore these questions in greater detail. If companies are allowed to claim the majority of the economic benefits from workers' data without adequately compensating workers, only to later displace those same workers who produced the data, it could lead to a situation where the value of human labor is systematically eroded in the long-term. This could have broader societal implications, including the exacerbation of income inequality and the undermining of both authors' and workers' rights in the digital economy. Yet if workers are allowed to claim significant, further remuneration for the use of their data, it could slow the deployment of new AI technologies in industries that need them. Workers' data is poised to become one of the most valuable corporate assets of the coming decade, and so if companies lack reliable access to such data, innovation could effectively stall. The challenge, therefore, will be to strike a balance between fairness and competitiveness, ensuring that corporate automation advances without systematically undervaluing the human contributions that make it possible.