

AI Revolution on Trial: Protection of Intellectual Property Rights against Generative Artificial Intelligence in the USA and UK

By Muhammad Ali Singay

Submitted to Central European University Department of Legal Studies

In partial fulfilment of the requirements for the degree of LL.M. in Global Business Law and Regulation

Supervisor: Prof. Marco Botta

Vienna, Austria

Central European University - Private University

ABSTRACT

The advent of Generative Artificial Intelligence ('Gen-AI') heralds a transformative era with the potential to reshape human society as we know it. In the global race for AI dominance, prominent programs like Chat GPT, Dall-E, and Stability Diffusion showcase the power of this technology. In the midst of this technological breakthrough, countries find themselves in a tug of war between their desire to dominate Gen-AI industries and to safeguard the Intellectual Property rights of the original creators.

The Gen-AI programs are trained on extensive datasets which includes writings, paintings, audios, videos and photographs, referred to as training datasets. These training datasets are primarily compiled through unauthorized data scraping from online sources which may also include copyrighted materials, raising ethical and legal concerns. This thesis specifically explores the legal and economic implications of the usage of copyrighted materials by the Gen-AI companies to train their models without the creator's permission or compensation. The paper aims to strike a balance between the protection of the rights of copyright holders and fostering innovation through Gen-AI development. This thesis initiates with the exploration of the legal and economic rationale behind copyright protection. The thesis further examines what are the Gen-AI models, how they are trained and how their training process infringes the copyright holders. The thesis also shed light on unauthorized training of Gen-AI models assessing its legal, ethical and economic impacts on creators, users, Gen-AI companies and on countries in general. This thesis then delves deeper

i

into the potential legal defenses for Gen-AI companies against the claims of copyright infringement, particularly the principle of 'fair use', 'fair dealing', 'text and data mining exceptions' and its interpretation under the copyright laws of the United States of America and the United Kingdom. Furthermore, the thesis explores the legal and economic dilemmas faced by the nations like the United States of America and the United Kingdom which are competing to be the forebears of the Gen-AI revolution. The training of Gen-AI programs on the datasets containing copyrighted materials falls under the category of copyright infringement under the laws of the United States of America and the United Kingdom. However, legislation of United Kingdom is planning to revise its copyright laws to facilitate the growth of Gen-AI companies despite of facing serious criticism from the copyright holders. On the other hand, the existing laws of the United States of America are vague on whether using copyrighted materials to train the Gen-AI programs falls under the existing exceptions of copyrights. There is extensive litigation going on the same question in courts of both jurisdictions. The last part of the thesis consists of the analysis of the potential solutions to balance between copyright protection and innovation in the domain of Artificial Intelligence including establishment of clearer legal frameworks and guidelines, collective licensing and royalties, implementation of technological solutions and international cooperation for protection of copyright. The final chapter synthesizes the discussion, draws conclusions and outlines a predictable way forward.

INTRODUCTION:	1
CHAPTER-I:	5
THE LANDSCAPE OF IP LAW & ARTIFICIAL INTELLIGENCE	5
1.1 Concept of Intellectual Property: A Friend or Foe of Creativity	5
1.2 Training Gen-AI Models: Methodology & Copyright Infringement	7
1.3 Competitive Dynamics & Copyright Implications in Gen-AI Development	10
CHAPTER-II:	
A COMPARATIVE ANALYSIS OF LEGAL AND ECONOMIC IMPLIC	ATIONS OF
COPYRIGHT LAWS IN USA AND UK	
2.1 Analysis of US Copyright Laws & Its Implication in Training Gen-AI Models	12
2.2 US: Potential Legal Defenses Against Copyright Infringement	15
2.3 Analysis of UK Copyright Laws & Its Implication in Training Gen-AI Models	21
2.4 UK: Potential Defenses Against Copyright Infringement	24
2.5 Adapting Copyright Laws for Gen-AI: A Way Forward for The USA & The UK	27
CHAPTER-III:	
NAVIGATING COPYRIGHT CHALLENGES: SOLUTIONS &	
RECOMMENDATIONS	
3.1 Establishment of Clearer Legal Framework & Guidelines:	31
3.2 Collective Licensing & Royalties	32
3.3 Implementation of Technological Solutions	34
3.3.1 Digital Watermarking	
3.3.2 Digital Rights Management (DRM)	
3.4 International Cooperation for Protection of Copyright	37
CONCLUSION:	
BIBLIOGRAPHY	

INTRODUCTION:

Generative Artificial Intelligence (Gen-AI) describes algorithms that can generate novel and diverse content, such as text, images, music, code, videos and more, based on a given input, termed as prompt¹, provided by the users. Gen-AI has been advancing rapidly in recent years, due to the development of powerful machine learning models, deep learning models and the availability of massive amounts of data. Some of the prominent examples of Gen-AI programs include Chat GPT, Dall-E, and Stability Diffusion, which can generate realistic and creative text, images, and videos, respectively.

The rise of Gen-AI has far-reaching implications across various aspects of life and industries, including entertainment, education, health and business. Gen-AI has a great potential to enhance human creativity, productivity, and innovation by offering new tools, insights, and solutions. The market research indicates that Gen-AI applications have the potential to add annually up to \$ 4.4 trillion to the global economy². This potential is clearly reflected in the rapid adaption rates of Artificial Intelligence which has increased to more than double over the past 5 years³. However, Gen-AI also poses significant challenges and risks, especially in the realm of intellectual property (IP) rights. One major source of contention is the use of

¹ Melvin Wong and others, "Prompt Evolution for Generative AI: A Classifier-Guided Approach." (*arXiv.org*, 24 May 2023) < https://arxiv.org/abs/2305.16347 > accessed 09 Apr. 2024.

² McKinsey & Company, "What Is Generative Ai?" (*McKinsey & Company*, 2 April 2024) http://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai accessed 09 Apr. 2024.

³ Michael Chiu and others, "The State of AI in 2022-and a Half Decade in Review." (*McKinsey & Company*, 6 December 2022) <<u>http://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review> accessed 12 Apr. 2024.</u>

copyrighted materials by the AI companies to train their Gen-AI models without the permission of the copyright holder and without providing any compensation.

Data is the gold mine for Gen-AI companies as they rely on extensive and diverse dataset to train and test their Gen-AI models. The datasets are often compiled through data scraping from various online sources including websites, blogs and social media platforms. These datasets may contain a variety of content, such as writings, paintings, photographs, songs, movies, and more, all of which may be protected under IP law like copyright, trademark and patent. This paper will exclusively focus on the copyright aspect of IP law in training the Gen-AI models and evaluate the legal, ethical and economic impact of copyright infringement on the creators, users, Gen-AI companies and countries as whole. The evidence presented in subsequent chapters of this paper indicates that these Gen-AI companies are not seeking consent from the copyright holder or paying the necessary royalties to them, thereby infringing upon the copyrights of the copyright holders.

This paper takes a comparative and analytical approach, focusing on the legal frameworks of the United States of America (USA) and the United Kingdom (UK), as both are at the forefront of Gen-AI innovations. Furthermore, the methodology implemented for this paper involves a comprehensive literature review, analyzing both academic and legal sources on training methods of Gen-AI models, copyright concepts, and infringement issues within the legal jurisdictions of the USA and the UK. Particular attention is given to the existing statutes, relevant case laws, Congress reports, UK Government consultation reports, and academic articles related to the topic. This detailed analysis aims to unveil any strengths and

2

weaknesses in the current legal structures, shedding light on challenges in addressing copyright infringement by Gen-AI models and governmental inclinations to favor Gen-AI companies at the expense of copyright holders, allowing the use of copyrighted materials without authors' permission.

In navigating the research, certain limitations emerge due to the dynamic nature of Gen-AI development, the multi-jurisdictional impacts of copyright laws and the lack of case presidents and a clear legal framework on the issue of training Gen-AI models on unauthorized copyrighted material. These limitations are thoroughly acknowledged and conscientiously examined within the research, providing a transparent framework for interpreting the findings and conclusions.

The structure of the paper is as follows:

Chapter I examines the rationale behind copyright protection, which serves to incentivize creativity by enabling authors and creators to gain monetary benefits from their work. However, we also take into account the criticisms of IP rights, including concerns about restricting the free flow of information and stifling creativity. The section also explores the delicate balance between safeguarding IP rights and promoting AI innovation, contrasting with the strategies employed by the USA and UK in this regard. Additionally, it presents an overview of the characteristics and evolution of Artificial Intelligence from Artificial Narrow Intelligence to Gen-AI. It also highlights the significance of data quality in Gen-AI training and the potential legal ramifications of using copyrighted material without the consent of the original author.

3

Chapter II analyzes the comparison between copyright laws in the USA and the UK and highlights key differences in their approaches and how these differences affect interpretations and enforcement of the copyrights. This chapter also examines how the training of Gen-AI models infringe the IP rights of the original creators, and the potential legal defenses such as fair use, fair dealing or text and data mining exemptions that they may invoke. We also assess the economic impact of these exceptions on the creators, users, and Gen-AI companies, in terms of incentives, welfare, and competition. Additionally, explores the national perspectives and policies of the USA and the UK, and how they are striving to strike a balance between protection of IP rights and the promotion of innovation in the domain of Gen-AI.

Chapter III proposes and evaluates potential solutions and recommendations for navigating the copyright challenges posed by the training of Gen-AI models. It considers various mechanisms, such as clearer legal framework and guidelines, collective licensing and royalties, technological solutions and international cooperation to protect and enforce copyright laws.

The paper intends to contribute to the academic and practical knowledge and understanding of the complex and dynamic relationship between IP laws and Gen-AI. The paper also hopes to stimulate further discussion and debate on this topic and to provide useful insights and guidance for the relevant stakeholders, such as policymakers, regulators, industry players, researchers, and users.

CEU eTD Collection

CHAPTER-I:

THE LANDSCAPE OF IP LAW & ARTIFICIAL INTELLIGENCE

1.1 Concept of Intellectual Property: A Friend or Foe of Creativity

The issue of copyright infringement in training of Gen-AI models is multifaced and unprecedented, hence, the legal framework is gradually evolving to address these challenges. In the absence of any clear precedence and guidelines, it is essential to start our journey from the very beginning to understand the origin and rational of copyright before delving into the complexities of copyright infringement in training Gen-AI models.

The term Intellectual Property (IP) is defined as "*a commercially valuable product* of human intellect, in concrete or abstract form such as copyrightable work, a protectable trademark, a patentable invention, or a trade secret"⁴. The underlying principle of IP protection is to grant creators exclusive rights to their creations and enable them to control and profit from their creative work. This framework incentivizes innovation and creativity by ensuring that creators can reap the rewards of their efforts. The rationale for IP rights can be traced back to economic and moral theories. Economical rationality dictates that by providing temporary exclusive rights to the authors for their creative work, IP rights enable the authors to secure financial returns on their intellectual creations and ensures that socially optimal amount of work is performed.⁵ The prospect of economic gain motivates individuals

⁴ Bryan A. Garner, *Black's Law Dictionary* (11th edn, Thomson Reuters 2019), 110.

⁵ Steven Shavell, *Foundations of Economic Analysis of Law* (Harvard University Press, 2004), 8.

and companies to invest time and resources into developing new ideas, technologies, and creative works leading to a dynamic and progressive society where innovation thrives. On the moral perspective, IP rights are grounded in the concept of natural rights, where creators are deemed to have an inherent right to control the products of their labor.⁶ The individuals who invest their labor, creativity, and skill into creating something new deserve to have their ownership rights recognized and protected. In this way the protection of IP rights also aligns with the broader principles of justice and fairness.

The protection of IP rights also has its critics. Some argue that IP laws can create barriers to entry and hinder the free flow of information and creates scarcity which inversely effects creativity and cumulative economic output.⁷ Critics claim that excessive IP protection can lead to monopolies, restrict access to knowledge, and stifle competition. They point to the fact that many great innovations have occurred in environments where information and ideas were freely exchanged. Furthermore, the exceptions of IP right granted by laws they are insufficient in present era. For example, copyright exceptions such as right to quote, make parody, use of work for educational purposes⁸ and principle of fair use⁹ are seen as inadequate for fostering innovation. The provisions like protection of copyrights up to 70 years after the death of the author or creator is more harming the creativity than favoring it.¹⁰ These concerns are amplified as Artificial Intelligent systems as they require vast amounts

⁶ Lawrence C. Becker, "The Labor Theory of Property Acquisition" (1976) 73 The Journal of Philosophy 653.

⁷ Daniel Pinheiro Astone, "Scarcity, Property Rights, Irresponsibility: How Intellectual Property Deals with Neglected Tropical Diseases" (2022) 34 Law and Critique, https://link.springer.com/article/10.1007/s10978-022-09324-3> accessed 12 Apr. 2024.

⁸ Simone Schroff "The Purpose of Copyright-Moving beyond the Theory" (2021) 16 Journal of Intellectual Property Law & Practice 16 (11) Journal of Intellectual Property Law & Practice https://doi.org/10.1093/jiplp/jpab130> accessed 12 Apr. 2024. ⁹ § 107, 17 U.S Code.

¹⁰ Jessica Silbey, *Against Progress: Intellectual Property & Fundamental Values in the Internet Age* (Stanford University Press, 2022), 21.

of data, often sourced from copyrighted materials, to learn and generate new content.

The pundits of modern economy are predicting that Artificial Intelligence has the potential to contribute up to \$15.7 trillion to the global economy by 2030.¹¹ This impact will be felt across various industries including healthcare, automotive, financial services, transportation, logistics, technology, communication, entertainment, retail, energy and manufacturing. However, the rapid growth of Artificial Intelligence also poses challenges related to protection of IP right. The governments and legislatures across the world are in tug of war between protection of the IP rights of the creators and to facilitate the rapid growth of Artificial Intelligence to reap its benefits. This tension is particularly evident in the approaches of the USA and the UK as they take different approaches to copyright protection. Both jurisdictions need to continuously adapt their legal frameworks to strike a balance that promotes both protection of IP rights and technological innovation. As the landscape of Artificial Intelligence and copyright law evolves, ongoing dialogue and cooperation between stakeholders, including governments, creators, and AIdevelopers, will be essential to achieving this equilibrium.

1.2 Training Gen-AI Models: Methodology & Copyright Infringement

Artificial Intelligence (AI) has become a widely recognized term. According to IBM report, 35% of companies are currently using AI in their business operations and an additional 42% are actively exploring the integration of AI into their businesses¹².

¹¹ Anand Rao and others, "Sizing the Prize-What's the Real Value of AI for Your Business and How Can You Capitalize?" (2018) 4 PwC's technical report on AI 1.

¹²IBM, and Morning Consult, "IBM Global AI Adoption Index 2022." (*IBM*, May 2022) https://www.ibm.com/downloads/cas/GVAGA3JP accessed 12 Apr. 2024.

AI is often described as a computer system that demonstrate human-like intelligence and cognitive abilities, such as deduction, pattern recognition, and the interpretation of complex data¹³. Over the period of time, AI has evolved from its humble beginning as Artificial Narrow Intelligence focused on menial, repetitive task to Generative Artificial Intelligence. Gen-AI utilizes machine learning and deep learning to enable software with human-like capacities to generate novel and diverse content including text, images, music, code and videos, based on the prompt¹⁴ of the users. It is pertinent to note that the field of Gen-AI is rapidly evolving and the new techniques and methodologies are continuously introduced to train these algorithms. The training of algorithms such as those used in machine learning and deep learning, involves complex and resources intensive processes including data collection and data processing. Machine learning can be thought of as an extension of predictive analytics, where algorithms detect and predicts data patterns, experiences and observations and automatically improves themselves¹⁵. The data that is provided to these algorithms are typically pre-processed, structured and labeled. While machine learning models can also be trained on unstructured data, however, the unstructured data must be first processed to convert into structured format.¹⁶ On the other hand, deep learning involves using unstructured data, where the algorithms "mimic the neural structure of the brain using learning, memory, and generalization. As a result, neural networks can capture highly complex relationships in the data and are used as a building block of sophisticated machine learning system (often called deep

¹³ Sanjiv Jaggia and others, *Business Analytics: Communicating with Numbers* (2nd edn, McGraw Hill 2023), 16. ¹⁴ Wong (n.1)

¹⁵ Angelo Kinicki and Denise Breaux Soignet, *Management: A Practical Introduction* (10th edn, McGraw Hill 2022), 31.

¹⁶ IBM, "What Is Strong AI?" (*IBM*, 13 October 2021) <http://www.ibm.com/topics/strong-ai> accessed 15 Apr. 2024.

learning)^{"17}. In Gen-AI the machine learning models not only utilize the capacity of prediction but also are able to generate new data.¹⁸

It is essential to gather a significant amount of data to train and test the Gen-AI algorithms. This data is downloaded on the servers, processed (including labeling or annotating), a model architecture is chosen and then the model is trained pursuing machine learning algorithm's with annotated data for a considerable amount of time.¹⁹ Once the model is trained, it is then tested with new data to further fine tune it.²⁰ Depending upon the Gen-AI model, the data may include text, images, audio, video or any other type of content relevant to the task the Gen-AI model is being trained to perform. The quality, diversity and proper processing of the collected data are crucial for the creation of a high-quality Gen-AI models. Advance Gen-AI software like Chat-GPT and Dall-E are trained on high quality data which are mostly, if not always, are copyright protected. In data processing, the collected data cleaned formatted properly which may involve removing irrelevant is and information, handling missing values, normalizing data, and converting it into a format that can be used by the AI algorithms²¹. The action of downloading (reproduction of data) the entire work or substantial portion of work and processing it to train Gen-AI algorithms without the permission of the original authors and under the ambit of copyright infringement.²² There are some creators falls

¹⁷ Jaggia and others (n.13)

¹⁸ Adam Zewe, "Explained: Generative AI" (*MIT News*, 9 November 2023) https://news.mit.edu/2023/explained-generative-ai-1109>accessed 18 Apr. 2024.

¹⁹ Microsoft Learn "Deep Learning vs. Machine Learning - Azure Machine Learning." (*Microsoft Learn*, 19 January 19 2024) accessed 18 Apr. 2024">https://learn.microsoft.com/en-us/azure/machine-learning/concept-deep-learning-vs-machine-learning?view=azureml-api-2>accessed 18 Apr. 2024.

²⁰ Zaid Alissa Almaliki "Do You Know How to Choose the Right Machine Learning Algorithm among 7 Different Types?" (*Towards Data Science*, 21 October 2022) < https://towardsdatascience.com/do-you-know-how-to-choose-the-right-machine-learning-algorithm-among-7-different-types-295d0b0c7f60> accessed 21 Apr. 2024.

²¹ Goyle Kartikay, Quin Xie, and Vakul Goyle. "DataAssist: A Machine Learning Approach to Data Cleaning and Preparation." (*Arxiv*, , July 2023) < https://arxiv.org/pdf/2201.07119> accessed 21 Apr. 2024.

²² § 106 Title 17 US Code, Section 16 Section Copyright, Designs and Patents Act 1988.

exceptions to the reproduction right of copyrighted material, which will be discussed in detail in later chapters, but those depends upon case-to-case basis.

1.3 Competitive Dynamics & Copyright Implications in Gen-AI Development

Data has become an incredibly valuable resource in present age of AI and the access to quality datasets for training Gen-AI models has a significant impact on the competitive dynamics within the industry. The large technology companies have the capital and resources to legally access a wide range of high-quality copyrighted data through licensing agreements or priority data collection methods. However, they often resort to illegally scraping data from internet to train their algorithms as a cheaper alternative, unjustly exploiting the copyright holders. The situation becomes more dismal in absence of clear guidelines on the legality of training AI models as these tech giants have strong financial and legal recourses to navigate complex copyright laws more easily. Furthermore, these companies are not only capitalizing on the scraped data to train their Gen-AI models but also creating a new market and ecosystems of AI as a service (AIaaS). Companies like Microsoft and Amazon Web Service employing concepts like transfer learning wherein the AI models that are trained on general datasets, for example large scale image datasets, and are packaged as pre-trained AI that can be employed for specific tasks by using a considerably smaller dataset that is problem specific.²³ The transfer learning enabled these companies to offer cloud AI applications, AI platforms and AI infrastructure. These tech giants are defending their actions by claiming to reduce customer's cost in

 ²³ Rudin Cynthia, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." (*Nature Machine Intelligence*, 13 May 2019)
 https://www.nature.com/articles/s42256-019-0048-x> accessed 18 Apr. 2024.

developing their own AI models as they can purchase pre-trained AI models and integrate them in their business. However, the provision of AIaaS further highlights the issue of unauthorized use of data specially copyrighted data for commercial purposes.

In the AI developer community, particularly among small AI developers, there is a growing demand for broader interpretation of copyright protection in the digital realm. It is essential to create a level playing field for AI developers in order to encourage innovation and competition within the market. If the copyright protections are strictly enforced, it will skew the competitive landscape against the small AI developers due to unequal access to data and resources. Even with innovative ideas, their growth is limited by financial and legal constrains. This action in long run will result into monopoly of few major companies in AI industry technological breakthrough and fewer opportunities for by smaller entities. Furthermore, the copyright holders would also be more inclined towards the big tech companies as they would be able to negotiate favorable licensing terms with them ensuring steady revenue stream for themselves as compared to negotiating licensing terms with small AI developers. Therefore, it is crucial to keep in mind the perspective of small AI developers when aiming to strike a balance between copyright protection and innovation.

CHAPTER-II:

A COMPARATIVE ANALYSIS OF LEGAL AND ECONOMIC IMPLICATIONS OF COPYRIGHT LAWS IN USA AND UK

2.1 Analysis of US Copyright Laws & Its Implication in Training Gen-AI Models

The founding fathers of United States recognized the importance of intellectual property and enshrined its protection in the Constitution of United States of America by granting Congress the power to "promote the Progress of Science and useful Arts, by securing for a limited Time to Authors and Inventors the exclusive Right to their respective Writings and Discoveries."24 Copyright is a form of intellectual property rights and is categorized as negative right as it imposes obligation on others not to infringe or interfere with the exclusive rights of copyright holder. In other words, copyright infringement is a strict liability offense which does not require the unlawful intent or culpability of the infringer/user and the infringer/user does not have to share copyrighted works in order to infringe a copyright²⁵. In USA, the primary statue that governs copyrights is Copyright Act of 1976 and codified in Title 17 of the United States Code. Under Copyright Act of 1976, copyright is defined as a set of exclusive rights granted to the authors for their original works of authorship including literary, musical, dramatic pictorial, sound recordings 26 . Furthermore, the author holds the exclusive right to reproduce, prepare derivatives,

²⁴ Art. I, § 8, cl. 8 United States Constitution.

²⁵ EMI Christian Music Grp., Inc. Vs. MP3Tunes, LLC 844 F.3d 79, 89 (2d Cir. 2016)

²⁶ § 102 Title 17 US Code

perform and display copyrighted work²⁷ as well as to transfer ownership rights of the copyrighted work by sale, rental, lease and lending²⁸. Any violation of the aforementioned exclusive rights of the author is categorized as infringement of copyrights under Section 501 of Copyright Act 1976.

Copyright law in USA is often referred to as providing a 'temporary monopoly' to the creator/author over the use, distribute and profit from their original work.²⁹ The emphasis on the economic basis of copyright, highlighting the temporary nature of the rights granted. The author's 'temporary monopoly' over their work extends throughout the life time of the author and expires 70 years after their death.³⁰ Once the copyright term expires, the work enters into the public domain and anyone can use them freely. However, in the case of copyright infringement, the author or copyright holder can sue the alleged infringers and is entitled to recover actual damages suffered along with the profits gained by the alleged infringer or the can claim statutory damages up to \$ 150,000 for willful infringement.³¹ These high fines act as a deterrent against copyright infringement.

It is also pertinent to understand the concept of 'derivative works' as such work is also used to train Gen-AI models. A derivative work is defined as 'work based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted'.³² This definition is broad and includes a wide range

 $^{^{\}rm 27}$ § 106 Title 17 US Code.

²⁸ § 106 Title 17 US Code.

²⁹ Georgia, et al. vs. Public. Resource. Org, Inc., 590 U.S. (2020)

³⁰ § 302, Tile 17 U.S. Code

³¹ § 504, Tile 17 U.S. Code.

³² § 101 Title 17 US Code.

of potential derivative works. Derivative works can be copyright and it extends only to the material contributed by the author of such work that clearly distinguished from the preexisting original work that is used in the derivative work.³³ The creator of a derivative work generally needs the permission of the copyright owner of the original work to create and distribute the derivative work legally.³⁴

The strict liability nature of copyright laws in USA makes it essential for the AI developers to ensure that they have necessary permissions to use copyrighted materials in their training dataset. Any action of unauthorized copying the entire work or substantial portion of work on a fixed medium and processing it to train Gen-AI algorithms falls under the ambit of copyright infringement³⁵. To fully grasp the concept of copyright infringement in this context, we need to understand the meaning of 'copies' and 'fixed' defined under the law. The Copyright Act of 1976 defines 'copies' as "material objects....in which a work is fixed by any method now known or later developed, and from which the work can be perceived, reproduced or otherwise communicated, either directly or with the aid of machine or device."36 Furthermore, a work is considered "fixed" in a medium when its copy is "sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for a period of more than transitory duration."37 The caselaw clearly categorize the storage of copies on permanent or temporary memory storage devices as a 'fixed' under copyright laws. In Stern Electronics, Inc vs. Kaufman³⁸, the Second Circuit Court held that the permanent storage of audiovisual

³³ § 103(b) Title 17 US Code.

³⁴ § 106(2) Title 17 US Code.

³⁵ § 106 Title 17 US Code.

³⁶ § 101, Tile 17 U.S Code.

³⁷ § 101, Tile 17 U.S Code.

³⁸ Stern Electronics, Inc vs. Kaufman, 669 F.2d 852 (2d Cir. 1982).

works of a videogame on a memory storage device are protected under the copyright law.³⁹ Similarly, in **MAI Systems Corp. vs. Peak Computer, Inc.**⁴⁰ the court addressing the issue of temporary storage of copyrighted work held that loading of copyrighted software in the random access memory (RAM) which temporarily stores the copyrighted work creates an unauthorize reproduction under Copyright Act of 1976⁴¹. Hence the permanent or temporary storage of copyrighted data on any devices by the AI developers to train their Gen-AI models falls under the category of making copies of the copyrighted work. However, there are exceptions to the copyright infringement under the law, such as fair use doctrine⁴² which may allow the use of copyrighted work without the need for permission of the author. The application of fair use doctrine requires the court to weight several statutory factors as discussed in detail below and is also strongly dependent upon the facts of the case.

2.2 US: Potential Legal Defenses Against Copyright Infringement

One of the arguable defenses for training Gen-AI models is that the copyrighted data used to train the Gen-AI models are taken merely for learning purposes to produce new work. It is analogous to a painter going to an art exhibition to get inspiration from the artist's work and to produce a new painting. In this analogy, the painting inspired by the artworks at the exhibition does not violate the copyrights of the artist who displayed their work. This raises the question that if a human can use previous work for inspiration then why a machine cannot use the previous work in the same

³⁹ ibid.

⁴⁰ MAI Systems Corp. v. Peak Computer, Inc., 991 F.2d 511 (9th Cir. 1993).

⁴¹ ibid.

⁴² § 107, Title 17 US Code.

way? The answer to this question lies in the method of inspiration. In the case of human beings, no copies of the previous work were made in any shape or form. However, to train a Gen-AI model it is necessary to make copies of the previous work to train it, hence the unauthorized copying of a copyrighted work infringes the copyright of the artist even though the copyrighted material is not reflected in the final work. **Sega Enterprises vs. Accolade, Inc.**⁴³ the court further explained this idea that the 'intermediate copying' of computer object code could constitute copyright infringement notwithstanding whether the end product of the copying also infringes those rights⁴⁴. In our above analogy, would it be copyright infringement if the human scanned the paintings by using his smartphone to use them later for his/her inspirational work? The answer is yes. The act of scanning the paintings and putting them on a digital storage device in preparation for further graphic insertion or its utilization as an inspiration constitutes copyright infringement as a matter of law.⁴⁵

The '**fair use doctrine**' can also be taken as a possible defense against the training of Gen-AI models on copyrighted data as it provides an exception for reproduction of copies of copyrighted work. Fair use doctrine stated under 17 U.S.C. § 107 varies from case to case basis depending on the four factors stated below:

- 1. the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
- 2. the nature of the copyrighted work;

⁴³ Sega Enterprises Ltd. v. Accolade, Inc., 785 F. Supp. 1392 (N.D. Cal. 1992).

⁴⁴ ibid.

⁴⁵ *Tiffany Design, Inc. v. Reno-Tahoe Specialty Inc.*, 55 F. Supp. 2d 1113 (D. Nev. 1999).

- 3. the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
- 4. the effect of the use upon the potential market for or value of the copyrighted work.

It is pertinent to note that at the time of writing this paper, there are no clear judgments by the court to determine the legality of training Gen-AI models on datasets containing copyrighted works. However, many lawsuits have been filed by the plaintiffs claiming infringement of the copyrights of their written and visual works by AI companies in training their AI models. These lawsuits include Authors Guild⁴⁶ and authors Paul Tremblay⁴⁷, Michael Chabon⁴⁸, Sarah Silverman⁴⁹, and others against Open AI; separate lawsuits by Michael Chabon⁵⁰, Sarah Silverman⁵¹, and others against Meta Platforms; proposed class action lawsuits against Alphabet Inc.⁵² and Stability AI and Mid journey⁵³; and a lawsuit by Getty Images against Stability AI.⁵⁴ In the Getty Images lawsuit, the plaintiff tried to address the plausible defense of fair use doctrine that might be taken by Stability AI stating that Stable Diffusion is a commercial product, weighing against fair use under the first statutory factor, and that the program undermines the market for the original works, weighing against fair use under the fourth factor.⁵⁵

⁴⁶ Authors Guild v. OpenAI Inc., 1:23-cv-08292, (S.D.N.Y.)

⁴⁷ Tremblay v. OpenAI, Inc., 3:23-cv-03223, (N.D. Cal.)

⁴⁸ Chabon v. OpenAI, Inc., 3:23-cv-04625, (N.D. Cal.)

⁴⁹ Silverman v. OpenAI, Inc., 3:23-cv-03416, (N.D. Cal.)

⁵⁰ Chabon v. Meta Platforms Inc., 3:23-cv-04663, (N.D. Cal.)

⁵¹ Kadrey v. Meta Platforms, Inc., 3:23-cv-03417, (N.D. Cal.)

⁵² L. v. Alphabet Inc., 3:23-cv-03440, (N.D. Cal. Jul 11, 2023) ECF No. 1

⁵³ Andersen v. Stability AI Ltd., 3:23-cv-00201, (N.D. Cal.)

⁵⁴ Getty Images (US), Inc. v. Stability AI, Inc., 1:23-cv-00135, (D. Del.)

⁵⁵ ibid.

The most relevant case to analyze the fair use doctrine regarding the training of Gen-AI model on datasets including copyrighted work is Sega Enterprises vs. Accolade, Inc.⁵⁶ In this case, Accolade reverse-engineered the video game programs of Sega to make their games compatible with Sega consoles. To achieve this result, Accolade copied Sega's copyrighted code in its entirety and then disassembled it to understand its workings. Afterward, Accolade created thier own games that were compatible with the Sega console but did not use any of Sega's code in its final product. Sega sued Accolade for copyright infringement. One of the important factors that the applet court analyzed was whether the disassembly of copyrighted code falls under the ambit of fair use doctrine? The Court held the actions of Accolade were justified under the fair use doctrine as per the factors described in 17 U.S.C. § 106. The first factor, "the purpose and character of the use," the court found that Accolade's purpose for reverse-engineering Sega code was essential to create games compatible with Sega's console which is a "legitimate, essentially non-exploitative purpose". Under the second factor, "the nature of the copyrighted work," Sega's video game code which contains functional aspects of console-compatibility, warranted a lower degree of copyrighted protection in comparison with traditional literary works. This analysis reflected the idea-expression dichotomy in the copyright law. acknowledging that ideas and functional concepts are not copyrightable. Under the third factor regarding the amount and substantiality of the copied portion, the court held that the copying of the entire code was insufficient to outweigh the favorable effects of the other three factors. Additionally, the court considered the fourth factor, "the effect of the use on the market", which favored Accolade as its entry into the

⁵⁶ Sega Enterprises Ltd. v. Accolade, Inc., 785 F. Supp. 1392 (N.D. Cal. 1992)

market for Sega's console compatible games would encourage creative expression in video games and promote competition in the market.

Another potential defense for training Gen-AI modes is the 'doctrine of transformative use' which the courts have used previously to balance between the technological growth and protection of IP rights of the authors/creators. The IP infringement cases that have argued transformative use as their defense, 94% of them were successful to come under the ambit of the fair use doctrine.57 Transformative use is where "the secondary use adds value to the original, if the quoted matter is used as a raw material, transformed in the creation of new information, new aesthetics, new insights and understandings.⁵⁸ For example, the Gen-AI models for self-driving cars are trained on a large dataset of images to recognize its surrounding can be categorized as transformative use. In Perfect 10, Inc. vs. Amazoon.com⁵⁹ the court held that the use of copyrighted pictures as thumbnails to showcase search engine results constitutes as transformative use and comes under the exception of the fair use. The images are originally intended for "entertainment, aesthetic, or informative function[s]," but are transformed by the search engine into a "pointer directing a user to a source of information".⁶⁰ Similarly in Authors Guild, Inc. vs. Google, Inc.⁶¹ the court held that the digitalization of copyrighted text by Google constitutes as fair use because the Google Book search engine was deemed highly transformative.⁶² Despite scanning complete copyrighted texts, Google only displayed snippets of text, effectively serving as pointers guiding

⁵⁷ Liu Jiarui, "An Empirical Study of Transformative Use in Copyright Law" (2019) 22 STAN. TECH. L. REV. 163, 174 < https://law.stanford.edu/wp-content/uploads/2019/02/Liu_20190203.pdf > accessed 12 April 2024.

⁵⁸ William Fisher, "Reconstructing the Fair Use Doctrine" (1988) 101 HARV.L.REV 1768.

⁵⁹ Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146 (9th Cir., 2007).

⁶⁰ ibid.

⁶¹ Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015)

⁶² ibid.

users to a wide array of books.⁶³ These instances highlight the trend of courts recognizing fair use exceptions to copyright infringement, particularly when the resulting product serves a functional purpose and offers greater social utility. If we follow the same logic in training Gen-AI models for surgery assistant robots on a large number of images, videos, audio and text can be reproduced and utilized as 'raw material' to teach the machine to make decisions like humans in particular categorized as creating "new insights and understanding". settings can be Furthermore, we can also deduce from the above cases that the commercial nature of a product does not affect its transformative value.

One should be cognizant of the fact the implementation of fair use doctrine to train the Gen-AI models is not straightforward rather it depends upon case-to-case basis. In **Fox News Network, LLC vs. TV Eyes, Inc.**⁶⁴ the Second Circuit court held that a TV clips search engine, while somewhat transformative in nature, does not qualify for fair use⁶⁵. This decision was primarily based on the finding that the search engine negatively affects the potential market for Fox News to license its works. Furthermore, in the earlier example of training the Gen-AI models of surgery assistant robots or self-driving cars, the court could counter-argue that the use of the informational aspect of copyrighted photographs to train the Gen-AI model to identify the objects did not fundamentally change the original purpose of the work, which is to depict such objects.

⁶³ Authors Guild, Inc. v. Google Inc., 954 F. Supp. 2d 282, 291 (S.D.N.Y. 2013), aff'd, 804 F.3d at 202.

⁶⁴ Fox News Network, LLC vs. TV Eyes, Inc. 883 F.3d 169, 178–80 (2d Cir. 2018)

⁶⁵ ibid.

2.3 Analysis of UK Copyright Laws & Its Implication in Training Gen-AI Models

In the UK, the primary statute dealing with copyrights is the Copyright, Design & Patents Act (CDPA) 1988. The close analysis of the USA and UK copyright laws shows many similarities as share their common origin in the British Statue of Anna⁶⁶. However, unlike the USA copyright law, the UK approaches copyright as an 'intangible property right' and treats creations of mind as a form of property that can be owned, sold or licensed like any other form of property.⁶⁷ Under Section 1 of CDPA 1988 copyright is defined as a property right granted to the authors for their original works of authorship of artistic work, literary work, musical work, dramatic editions⁶⁸. work. sound recordings. typographical arrangement of published Furthermore, the author has the exclusive right to reproduce, prepare derivatives, perform and display copyrighted work in public.⁶⁹ The author further has the exclusive right to rental, lease and lend the copyrighted work to the third party as per the terms and conditions mutually decided between the parties.⁷⁰ The copyright owner enjoys the exclusive rights regarding its original creation for a limited period of time.⁷¹ For example, the Copyright term for literary, dramatic, musical or artistic work extends throughout the lifetime of the author and expires 70 years after the author's death⁷². After the copyright term expires, the works enter into the public domain and anyone can use them freely. However, during the copyrighted term a

⁶⁶ Sandra Day O'Connor, "Copyright Law from an American Perspective" (2002) 37 Irish Jurist https://www.jstor.org/stable/44027015 > accessed 12 April 2024.

⁶⁷ Hanger Holdings v Perlake Corporation SA & Anor [2021] EWHC 81 (Ch).

⁶⁸ Section 1, Copyright, Design & Patents Act 1988.

⁶⁹ Section 16, Copyright, Design & Patents Act 1988.

⁷⁰ Section 16 (1) (ba) and Section 18 (A) of Copyright, Design & patents Act 1988.

⁷¹ Designers Guild Ltd v Russell Williams (Textiles) Ltd [2000] 1 WLR 2416, [2001] UKHL 17.

⁷² Section 12 (2) Copyright, Design & patents Act 1988.

person may commit copyright infringement if he/she makes unauthorized copies of the copyrighted work or import copies of work that infringes copyright to the UK.⁷³

Under CDPA 1988 the concept of derivative work is similar to that in the USA, however, there is more emphasis on the requirement that the derivative work must itself be an original work of skill, labor and judgment.⁷⁴ The derivative work must be significantly different from the original work to be subject to copyright in its own right. In both jurisdictions, the creator of a derivative work generally needs the permission of the copyright owner of the original work to create and distribute the derivative work legally.⁷⁵

In case of infringement of copyright, the author or copyright holder can pursue civil remedies and/or criminal offenses against the alleged infringer(s). The civil remedies may include injunction, damages and an account of profit⁷⁶. Copyright infringement is also a strict liability offence under CDPA 1988. However, in case of unintentional infringement of copyright, the author will not be entitled to any damages but without prejudice to any other remedy available.⁷⁷ Under criminal offense upon conviction on indictment, the penalties include a fine and/or imprisonment of up to 10 years.⁷⁸ These high fines and penalties act as a deterrent against copyright infringement, however, the courts have considerable discretion for the interpretation of these offenses under CDPA 1988. There are exceptions to copyright infringement, such as doctrine of fair dealing which may allow use of

⁷³ Section 17 and Section 18 Copyright, Design & patents Act 1988.

 ⁷⁴ Copyright Witness Ltd, "A Derivative Work Is a Work That Is Based on (Derived from) Another Work." (UK

 Copyright
 Service,
 11
 November
 2019)

 <https://copyrightservice.co.uk/copyright/p22_derivative_works?ssp=1&darkschemeovr=1&setlang=en&cc=AT</td>

 &safesearch=moderate> accessed 02 May 2024.

⁷⁵ Section 16, Copyright, Design & patents Act 1988.

⁷⁶ Section 96 Copyright, Design & Patents Act 1988.

⁷⁷ Section 97(1) Copyright, Design & Patents Act 1988.

⁷⁸ Section 107 (4A) (b) Copyright, Design & Patents Act 1988.

copyright materials without the need for permission under certain circumstances. However, the scope of the doctrine of fair dealing is restrictive in comparison with the doctrine of fair use. Under Section 29, CDPA 1988 the utilization of a copyrighted work is categorized as fair dealing if the use of copyrighted work is for the purpose of non-commercial research. Section 29(A), CDPA 1988 further states that the copying of a copyrighted work (text and data mining) 'to carry out a computational analysis of anything recorded in the work for the sole purpose of research of a non-commercial purpose' also falls under the exceptions of copyright infringement.⁷⁹

Now we have become acquainted with the basic structure of UK copyright laws, we can apply this knowledge in the context of training the algorithms of Gen-AI models. Dataset is one of the important pieces of the puzzle in training Gen-AI models and under CDPA 1988 datasets/databases are protected. The law defines database as "a collection of independent works, data or other materials which – (a) are arranged in a systematic or methodical way, and (b) are individually accessible by electronic or other means."⁸⁰. It is pertinent to distinguish that a data set may also consist of individual copyrighted materials and authorization to use the copyrighted materials does not automatically translate into authorization to use the data set. The entire collection of databases can be protected under copyright laws as literary work if "the selection or arrangement of the contents of the database represents the author's own intellectual creation". ⁸¹ We are aware that the dataset, which may also contain copyrighted work, is in whole or the substantial part is copied on the servers

⁷⁹ Section 29A Copyright, Design & Patents Act 1988.

⁸⁰ Section 3A Copyright, Design & Patents Act 1988.

⁸¹ Section 3A (2) Copyright, Design & Patents Act 1988.

of AI companies, potentially infringing the copyrights of the right holders.⁸² Furthermore, the copyrighted work and/or the data set which is compiled under the doctrine of fair dealing for academic research purposes cannot be transferred to any third party for non-commercial use as it falls under copyright infringement.⁸³ Additionally, when the Gen-AI model generates a work that reproduces in whole or a substantial part of the original copyrighted work that was used for training they may commit to copyrig and communication to the public which falls under the infringement of copyrights under Section 16(2) and Section 17 of CDPA 1988. As discussed above, the potential defense of fair dealing is also unlikely to apply in the case of training Gen-AI models as these models are trained by the companies for commercial purposes.

2.4 UK: Potential Defenses Against Copyright Infringement

One of the arguable defenses for training Gen-AI models on copyrighted data is the Gen-AI model makes transient copies of the copyrighted work which is an essential part of the technological training process. The act of training the Gen-AI model per se has less to do with economic significance and more to do with the development and operation of new technologies.⁸⁴ To understand the exception regarding the creation of temporary copies, we have to first understand Section 28 A of CDPA 1988 which was added to the CDPA 1988 via regulation in 2003 to give effect to Directive 2001/29/EC of 22 May 2001 ("Directive"). Article 5.1 of the Directive defines the criterion to qualify for the exemption of the act of reproducing/copying of copyrighted work without authorization of the copyright holder. However, the

⁸² Section 17 (2) of Copyright, Design & Patents Act 1988.

⁸³ Section 29A (2) Copyright, Design & Patents Act 1988.

⁸⁴ Section 28A Copyright, Design & Patents Act 1988.

exception under Article 5.1 of the Directive shall only be applied in "special cases which do not conflict with a normal exploitation of the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the rightsholder".⁸⁵ In **Infopaq International A/S v. Danske Dagblades Forening**⁸⁶, the Court of Justice of European Union ("CJEU") paraphrased Article 5.1 of the Directive and stated that the exception for the reproduction of copyrighted material is "dependent of the following five conditions being fulfilled:

- 1. the act is temporary;
- 2. it is transient or incidental;
- 3. it is an integral and essential part of a technological process;
- 4. the sole purpose of that process is to enable a transmission in a network between third parties by an intermediary of a lawful use of a work or protected subject-matter; and
- 5. the act has no independent economic significance."⁸⁷

The primary purpose of the Directive was to strike a balance between the rights of the copyright holders in the reproduction of their work and to facilitate the development and operation of new technologies and the rights of the users who wish to use those technologies.⁸⁸ In order to qualify as 'transient' the Gen-AI companies have to devise an automated mechanism, without human intervention, wherein the copied material used for training the machine learning algorithms should only be stored for a limited duration and automatically deleted as soon as the training ends.⁸⁹

⁸⁵ Article 5.5 Directive 2001/29/EC of 22 may 2001.

⁸⁶ (Case C-5/08) Infopaq International A/S v. Danske Dagblades Forening [2010] F.S.R. 495

⁸⁷ ibid.

⁸⁸ Football Association Premier League Ltd v QC Leisure and Karen Murphy v Media Protection Services Ltd [2012] 1 CMLR 769

⁸⁹ (Case C-5/08) Infopaq International A/S v. Danske Dagblades Forening [2010] F.S.R. 495

The requirement of the third condition that the copying should be an "integral and essential part of a technological process" meant that "the completion of the temporary reproduction is necessary, in that the technological process concerned could not function correctly and effectively without that act".90 In the case of training Gen-AI models the machine learning algorithms can only be trained by temporarily copying the datasets to transform them into digital data that can be read by the machine learning algorithms. The stance of CJEU regarding the fourth requirement is beautifully summarized by the court in Public Relations Consultant Association Limited vs. The Newspaper Licensing Agency Limited and others⁹¹. The court state that "the exception is not limited to copies made in order to enable the transmission of material through intermediaries in a network. It also applies to copies made for the sole purpose of enabling other uses, provided that these uses are lawful. These other uses include internet browsing.... for the purpose of article 5.1(Directive), a use of the material is lawful, whether or not the copyright owner has authorized it, if it is consistent with EU legislation governing the reproduction right, including article 5.1 itself.... The use of the material is not unlawful by reason only of the fact that it lacks the authorization of the copyright owner." ⁹² Under this comprehension of the court we can argue that the temporary copies of the work use to train the machine learning algorithms can fall under 'sole purpose of enabling other uses' and such use is also lawful in nature. The fifth requirement is the most controversial in nature when it comes to training the machine algorithms of Gen-AI model. In Football Association Premier League Ltd v QC Leisure⁹³ and Karen

⁹⁰ (Case C-302/10) Infopaq International A/S v. Danske Dagblades Forening ("Infopaq II"), 17 January 2012.

⁹¹ *Public Relations Consultant Association Limited vs. The Newspaper Licensing Agency Limited and others* [2013] UKSC 18.

⁹² ibid.

⁹³ (Case C-403/08) Football Association Premier League Ltd v QC Leisure and Others.

Murphy v Media Protection Services Ltd⁹⁴ and in Infopaq International A/S v. Danske Dagblades Forening⁹⁵ the CJEU further clarified the "independent economic significance" and stated that the requirement of making the temporary copies should not have any independent economic significance does not mean that these temporary copies should not have any commercial value. The court emphasizes that the economic advantage from making temporary copies should not be distinct from the advantage of using the original work. The large LLM models like Chat GPT actively decline the request for presentation of a copyrighted work. In simpler terms, the copies should not have their own market or use apart from helping users to access the original content. The court further stated that the temporary copies can be made to increase the efficiency of the process and should not create separate economic benefits. If we analyze the training process of Gen-AI models in light of the above criterion we can argue that the copies that are made to train the machine learning algorithms are temporary in nature and are not utilized as a substitute for the original work but only used to improve the performance of machine learning models. Furthermore, the value of these temporary copies in the training process is directly tied to training the algorithms and not been used to any other commercial use.

2.5 Adapting Copyright Laws for Gen-AI: A Way Forward for The USA & The UK

In the above comparative analysis of the legal frameworks in both jurisdictions, it is evident that both legal frameworks treat copyright differently and therefore require

⁹⁴ (Case C-429/08) Karen Murphy v Media Protection Services Ltd [2012] 1 CMLR 769.

^{95 (}Case C-302/10) Infopaq International A/S v. Danske Dagblades Forening, 17 January 2012.

different approaches to addressing this issue. In the USA, copyright law grants a 'temporary monopoly' to the creator/author over the use, distribution and profit from their original work.⁹⁶ The emphasis on monopoly protection in the USA underscores the economic rationale for copyright, highlighting the temporary nature of the rights granted. This economic rationale is one of the reasons for the broader and more flexible interpretation of the doctrine of fair use, allowing the courts to assess copyright infringement issues on a case-by-case basis. Many legal experts in the USA believe that the flexibility of the doctrine of fair use is sufficient to address the issue of copyright infringement in training Gen-AI models. However, there remains a need to establish clear guidelines on the legality of training Gen-AI models and ensuring fair compensation for copyright holders. It is also essential for the government to facilitate the ongoing dialogue between AI developers, copyright holders and legal experts to adapt the existing legal framework to every evolving technological landscape. Moreover, the US Congress is also actively monitoring the landscape of Gen-AI and copyright laws which indicates potential future legislative action to provide more definitive solutions.

On the other hand, the UK approaches copyright as an 'intangible property right,' considering intellectual creations as a form of property that can be owned, sold, or licensed like any other form of tangible property. This concept of tying copyright with property rights gives the right holder an inalienable right to their intellectual creations as per the CDPA 1988. Furthermore, the narrow and strict interpretation of the doctrine of fair dealing poses an additional hurdle to finding any leeway for the commercial training of Gen-AI models. However, courts could potentially broaden

⁹⁶ Georgia, et al. vs. Public. Resource. Org, Inc., 590 U.S. ___ (2020)

their interpretation of doctrine of fair dealing as discussed above and/or text and data mining exception in Section 28A of the CDPA 1988 and extend it as an exception for training Gen-AI models as well depending on case-to-case basis. Currently, the UK legislature needs to make amendments to existing laws to address the challenges posed by AI and strike a balance between copyright protection and innovation in AI. The UK legislature initially announced a proposal to extend text and data mining exception to training AI models, however, this proposal was later withdrawn by the government due to the pressure from copyright holders.⁹⁷ According to the statistical report by the UK Intellectual Property Office, the businesses in UK invest more than GBP 130 billion a year in knowledge assets and out of which GBP 63 billion are protected under IP rights.⁹⁸ UK legislature is taking proactive steps to address these challenges and has developed a National AI Strategy⁹⁹ and conducted multiple consultation sessions with the key stakeholders to formulate policies that can strike a balance between copyright protection and innovation.

⁹⁷ Rachel Montagnon, "UK Withdraws Plans for Broader Text and Data Mining (TDM) Copyright and Database Right Exception", (*Herbert Smith Freehills* / *Global law firm*, 1 March 2023) <https://www.herbertsmithfreehills.com/notes/ip/2023-03/uk-withdraws-plans-for-broader-text-and-data-mining-tdm-copyright-and-database-right-exception > accessed 04 Jun. 2024.

⁹⁸ Intellectual Property Office, and UK IPO, Intellectual Property Office § (2022). https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents/outcome/artificial-intelligence-and-intellectual-property-copyright-and-patents-government-response-to-

patents/outcome/artificial-intelligence-and-intellectual-property-copyright-and-patents-government-response-toconsultation. Accessed 04 Jun. 2024.

⁹⁹ Intellectual Property Office, and UK IPO, National AI Strategy § (2022). accessed">https://www.gov.uk/government/publications/national-ai-strategy/national-ai-strategy-html-version>accessed 02 May 2024.

CHAPTER-III:

<u>NAVIGATING COPYRIGHT CHALLENGES: SOLUTIONS &</u> <u>RECOMMENDATIONS</u>

We can observe that the rise of Gen-AI has sparked a mixture of hopes and concern about what future holds for us. It is abundantly clear that integration of AI in our existing structures is inevitable and represent the next stage in the evolution human society. However, this should not mean that we give a green flag to Gen-AI development and sacrifice the rights (copyrights) of the intellectuals who has brought us to this point. We see the nation states grappling to balance between their desire to dominate Gen-AI industries and to safeguard the Intellectual Property rights of the original authors/creators. The tension between protecting copyright holders and fostering innovation becomes more pronounced as AI technology continues to evolve. To address this complex issue, we need a multifaceted approach potentially consisting of clearer legal framework and guidelines, collective licensing royalties, technological solutions and international cooperation to and protect copyright. Additionally, establishing a regulatory sandbox could serve as a crucial testing ground for these solutions¹⁰⁰. A regulatory sandbox will offer a controlled aforementioned environment the practical implications of the to assess recommendations and to adjust the solutions to get the desired balance between IP right protection and fostering innovation. Furthermore, the regulatory sandbox will

¹⁰⁰ Wolf Georg Ringe, "Why We Need a Regulatory Sandbox for AI" (*Oxford Law Blogs*,12 May 2023) <https://blogs.law.ox.ac.uk/oblb/blog-post/2023/05/why-we-need-regulatory-sandbox-ai>accessed 11 May 2024.

also provide a forum to strengthen the collaboration between all stakeholders to find an amicable solution for training Gen-AI algorithms.¹⁰¹

3.1 Establishment of Clearer Legal Framework & Guidelines:

In both jurisdictions of the USA and UK there is a heated debate on the efficacy of existing IP laws to address the nuance challenges brought by AI. In USA the majority opinion rest with the idea that the fair use doctrine has the potential to address the issue of IP rights in training AI models. Conversely, the UK government is actively pursuing legislative changes to achieve its aim of making UK the global AI superpower. Establishing clear legal frameworks for training AI models is crucial to promote innovation in AI industry, as it encourages businesses to invest and expand in this sector. Therefore, the legislative reforms crucial to adapt copyright laws to the realities of AI technology. Legislatures in both the US and UK should develop a system or mechanism which enables the copyright holders to identify their work in the training dataset of Gen-AI algorithms and also define what constitutes acceptable use of copyrighted material in AI training for both commercial and noncommercial use. This could involve setting limits on the amount of data that can be used, the types of works that can be included, and the specific conditions under which training of AI models on copyrighted work is permissible. The government can issue guidelines defining the parameters of proportionality and necessity regarding the use of copyrighted work to train AI models. This principle will ensure

¹⁰¹ Lauren A Fahy, "Fostering regulator–innovator collaboration at the frontline: A case study of the UK's regulatory sandbox for fintech." (2022) 44 Law & Policy < https://doi.org/10.1111/lapo.12184 > accessed 12 May 2024.

that only the necessary amount of data is utilized to achieve the AI's purpose without unfairly exploiting the copyright holder's work.

The doctrine of 'fair use' in the USA and 'fair dealing' and 'text and data mining' in the UK should also be adapted to consider the non-commercial use of copyrighted material to train AI algorithms where there is a minimum impact on the market of copyrighted material. This adaptation can create an exception for the AI models that are not in direct competition with the market of copyrighted material. For instance, training the automatic driving algorithms using the dataset of pictures and videos may not unjustly exploit the copyrights of photographers and videographers and their potential earnings, as both are targeting different markets. Governments should also facilitate the ongoing dialogues between stakeholders, including copyright holders, AI developers, legal experts, and policymakers, to ensure that laws and policies keep pace with technological advancements and reflect the interests of all parties involved. The UK Government has developed a National AI Strategy¹⁰² and held multiple consultation sessions with the key stakeholders to formulate policies that can strike a balance between copyright protection and innovation. US Congress is also actively monitoring the landscape of AI and copyright laws.

3.2 Collective Licensing & Royalties

Licensing for the use of copyright materials is well established practice in the creative industry. Online platforms like Spotify and Netflix are few of the examples where the services provider negotiated with the copyright holders and acquire the

¹⁰²Intellectual Property Office, and UK IPO, National AI Strategy (2022). https://www.gov.uk/government/publications/national-ai-strategy/national-ai-strategy-html-version accessed 11 May 2024.

license of their work. The case of Gen-AI companies is not much different than these platforms and there is an essential need to establish a robust licensing framework to balance the interest of copyright holders and AI developers. Some platforms such as Getty Images "provides licenses to leading technology innovators for purposes related to training AI systems in a manner that respects personal and property rights".¹⁰³ These licenses intellectual types of are called 'limited licenses'¹⁰⁴ wherein the copyright holder only grants license regarding some specific and limited use of its work. This licensing tool not only solve the issue of legal uncertainty in case of training AI models but also ensures a fair compensation for the copyright holder for use of their work. The stakeholders can also consider establishing a standardized licensing agreement that can be managed through Collective Management Organizations (CMOs). While the establishment of CMOs can cause additional cost for the copyright holders, the government/regulator can assume this task to ensure the enforcement of such licensing agreements. CMOs can also streamline the process by granting licenses for large datasets used in training of AI models, with AI developers would pay a royalty fee/license fee to the CMO that is then distribute to the copyright holders. Such system ensures that creators benefit financially from the use of their works and simplifies the legal complexities for AI companies.

We can also introduce some variations to the licensing models to better cater a wide spectrum of AI models. For instance, a tiered licensing model can be considered,

¹⁰³Getty Images, "Getty Images Statement" (Getty Images Statement, 17 January 2023) https://newsroom.gettyimages.com/en/getty-images/getty-images-statement accessed 4 June 2023 ¹⁰⁴Oxford "Limited (Oxford Reference, License" *Reference*)<https://www.oxfordreference.com/display/10.1093/oi/authority.20110803100106217> accessed 11 May 2024.

where the licensing fee is determined based on the nature and scale of AI projects. This could mean that non-competing commercial and research-based AI initiatives may qualify for lower licensing fees or exemptions. This approach aims to promote innovation and discourage any unfair exploitation of copyright holders.

3.3 Implementation of Technological Solutions

The copyrighted work on the internet can be protected from data scraping by implementing technological solutions such as digital watermarking and digital rights management (DMR).

3.3.1 Digital Watermarking

Digital watermarking technology involves embedding information directly into digital content in a manner that is typically imperceptible to the human senses but can be easily detected and read by computers. This technique can be applied to images, videos, audio, or text and the imbedded information can be identity of copyright holder, usage rights, and other pertinent metadata. Unlike traditional watermarks that are visible and can be easily removed or tampered with, digital watermarks are embedded into the content's underlying data structure, making them much harder to detect and eliminate without degrading the quality of the content itself.

There are methods, such as spatial domain and frequency domain, that can be implemented to digitally watermark images, audio and videos. Spatial domain techniques directly alter the pixel values, which can be visible or invisible to the

CEU eTD Collection

34

naked eye, depending on the implementation.¹⁰⁵ On the other hand, frequency domain techniques, on the other hand, embed the watermark into the frequency components of the audio/video content, making it immune to common editing operations like resizing and compression.¹⁰⁶ These methods ensure that even if the content undergoes various transformations, the digital watermark remains intact and verifiable. In the context of text, digital watermarks can be embedded by altering text properties such as spacing, font, or by incorporating invisible characters that do not affect the readability of the text but can be detected by software tools.¹⁰⁷ These embedded watermarks act as a fingerprint that can be traced back to the original creator, providing irrefutable evidence of ownership. It will also act as a deterrent against unauthorized use of the work as the embedded watermark can provide the necessary evidence to pursue legal action against the infringers.

Furthermore, combining digital watermarking with technologies like blockchain creates a tamper-proof record of ownership and transaction history, enhancing reliability and trustworthiness in protecting copyrighted materials. Overall, digital watermarking provides a powerful tool for copyright holders to safeguard their works and ensure their intellectual property rights are respected in the digital landscape.

¹⁰⁵ Zihan Yuan and others "A blind image watermarking scheme combining spatial domain and frequency domain" (2020) 37 The Visual Computer, 1867–1881 https://doi.org/10.1007/s00371-020-01945-y, accessed 14 June 2024.

¹⁰⁶ Ibid

¹⁰⁷ Aiwei Liu and others "A Survey of Text Watermarking in the Era of Large Language Models" (2023) 1 (1) Arxiv < DOI:<u>10.48550/arXiv.2312.07913</u>> accessed: 14 June 2024.

3.3.2 Digital Rights Management (DRM)

Digital Rights Management (DRM) offers a comprehensive and effective framework for safeguarding copyrighted material by controlling the access, usage and distribution of digital content. DRM systems operate by encrypting digital content and implementing strict access controls to ensure that only authorized users can access the content under specified conditions set by the copyright holder¹⁰⁸. These conditions may include subscription-based access, one-time viewing permissions, or limits on the number of devices on which the content can be accessed. Such controls ensure that the content is only used in the manner intended by the copyright holder and also prevents unauthorized copying, sharing. and distribution. thereby safeguarding the rights of content creators and copyright holders. This granular control over usage helps to protect the economic value of the content, ensuring that creators are fairly compensated for their work. Monitoring and enforcement are critical components of DRM. They can track how content is used and detect patterns that indicate unauthorized usage. For example, if a single user account is used to access content from multiple devices in a suspicious manner, the DRM system can flag this for further investigation. This monitoring capability not only helps in detecting unauthorized use but also provides valuable insights into how the content is consumed, which can be used in identifying the targeted market for future licensing and distribution strategies.

DRM can be integrated with digital watermarking to create a robust protection system. While DRM controls access and usage, digital watermarking embeds

¹⁰⁸ Valentin Mulder and others *Trends in Data Protection and Encryption Technologies* (Springer 2023)

ownership information directly into the content. This combined approach ensures that even if the DRM controls are somehow bypassed, the watermark can still serve as evidence of ownership and trace unauthorized use back to the original source. This combined strategy not only protects the economic interests of the creators but also encourages continuation in innovation and creativity.

3.4 International Cooperation for Protection of Copyright

Internet has transcended traditional international boundaries of nation-states and it is widely acknowledged fact that a significant amount of data that is used to train Gen-AI algorithms are scraped from internet. This cross-border nature of internet undermines the effective enforcement of copyright laws especially in case of protecting digital content from unauthorized use, including for training AI algorithms. This serious situation calls for an enhance international cooperation in protecting the rights of copyright holders. We need to strengthen the existing frameworks and creating new mechanisms for collaboration. One potential approach could involve amending the international agreements, such as the Berne Convention the Agreement on Trade-Related Aspects of Intellectual Property Rights and (TRIPS), to include provisions specifically addressing the challenges posed by AI. These amendments could establish clearer guidelines for the use of copyrighted outline mechanisms material in AI training datasets and for cross-border enforcement of copyright laws. Additionally, the establishment of a global task force or working group dedicated to addressing copyright issues related to AI could facilitate information sharing, best practice exchange, and coordination among countries. This task force could work closely with existing organizations such as the World Intellectual Property Organization (WIPO) and the World Trade Organization

37

(WTO) to ensure that copyright protection remains a priority on the international agenda.

CONCLUSION:

In our research, we delved into the multifaceted issues of copyright law in relation to the training of Gen-AI models, with a specific focus on the legal frameworks of the USA and the UK. Through a detailed examination of the copyright laws of both jurisdictions, we uncovered the key challenges and deficiencies that arise when traditional copyright concepts are applied to the rapidly advancing field of AI. Our comparative analysis revealed that although the USA and the UK have common roots in copyright principles, they differ significantly in their approach. The USA's broad interpretation of fair use provides a flexible framework that could potentially accommodate the complexities of Gen-AI model training. However, the absence of clear statutory guidance and case law highlights the need for explicit legal standards to address the nuances of AI technology and ensure fair compensation for copyright holders. Conversely, the UK's treatment of copyright as an intangible property right, along with a narrower interpretation of fair dealing, presents substantial obstacles for the commercial training of Gen-AI models. The legal framework of the UK, as codified in the Copyright, Designs, and Patents Act 1988 (CDPA 1988), requires legislative amendments to strike a better balance between copyright protection and technological innovation. The recent withdrawal of the proposal to extend text and data mining exceptions underscores the tension between safeguarding copyright holders' rights and promoting AI development.

The training of Gen-AI models under current copyright laws raises significant concerns about potential infringement on the rights of copyright holders. Both the USA and the UK lack sufficient provisions or exceptions that justify the widespread use of copyrighted material in training AI models without explicit permission of the

39

author/creator. This copyright infringement in training of AI models stems from fundamental activities involved in reproducing and processing of copyrighted work without the consent of the author/creator. The potential defenses against copyright infringement like fair use in the USA and fair dealing in the UK have limited applicability due to their restrictive interpretations and the commercial nature of AI development. To overcome these challenges, it is crucial to carefully evaluate potential exceptions and reforms. Establishing clear legal guidelines that define the boundaries for using copyrighted materials in AI training can bring much-needed clarity and security for both creators and developers. Similarly, collective licensing schemes can facilitate fair compensation for the authors/creators while allowing the use of large datasets necessary for AI innovation. In addition, technological solutions such as digital watermarking and strengthened digital rights management help monitor and enforce copyright compliance (DRM) systems can more effectively.

In conclusion, the current legal frameworks pose significant challenges to the training of Gen-AI models, often resulting in clear copyright infringements under existing laws. Nevertheless, these challenges also offer a vital opportunity for thoughtful reform. As copyright law and AI technology continue to progress, it's crucial to maintain a balance between protecting rights of copyright holders and fostering technological innovation. The ongoing dialogue between the stakeholders, comprehensive research, and proactive involvement of legislature in this regard shows the determination of governments to achieve this equilibrium. Such equilibrium will not only safeguard the interests of copyright holders but also promote the responsible and ethical advancement of AI technologies. It is pertinent to remember that the future of copyright law in the AI era hinges on our ability to

40

establish a legal framework that encourages both creativity and technological progress. By exploring potential exceptions and reforms, we can better integrate AI into our legal and creative ecosystems. This effort will not only protect intellectual property but also foster innovation and growth, inspiring us to shape a future where copyright law effectively supports the transformative potential of AI.

BIBLIOGRAPHY

Almaliki Z, "Do You Know How to Choose the Right Machine Learning Algorithm among 7 Different Types?" (*Towards Data Science*, 21 October 2022) <https://towardsdatascience.com/do-you-know-how-to-choose-the-right-machine-learningalgorithm-among-7-different-types-295d0b0c7f60> accessed 21 Apr. 2024.

Andersen v. Stability AI Ltd., 3:23-cv-00201, (N.D. Cal.)

Astone DP, "Scarcity, Property Rights, Irresponsibility: How Intellectual Property Deals with Neglected Tropical Diseases" (2022) 2022) 34 Law and Critique https://link.springer.com/article/10.1007/s10978-022-09324-3 accessed 12 Apr. 2024.

Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015)

Authors Guild v. OpenAI Inc., 1:23-cv-08292, (S.D.N.Y.)

Authors Guild, Inc. v. Google Inc., 954 F. Supp. 2d 282, 291 (S.D.N.Y. 2013), aff'd, 804 F.3d at 202

Becker LC, "The Labor Theory of Property Acquisition" (1976) 76 The Journal of Philosophy 653.

Case C-302/10 Infopaq International A/S v. Danske Dagblades Forening ("Infopaq II"), 17 January 2012.

(Case C-403/08) Football Association Premier League Ltd v QC Leisure and Others.

(Case C-429/08) Karen Murphy v Media Protection Services Ltd [2012] 1 CMLR 769.

(Case C-5/08) Infopaq International A/S v. Danske Dagblades Forening [2010] F.S.R. 495.

Chabon v. Meta Platforms Inc., 3:23-cv-04663, (N.D. Cal.)

Chabon v. OpenAI, Inc., 3:23-cv-04625, (N.D. Cal.)

Chui M and others, "The State of AI in 2022-and a Half Decade in Review" (*McKinsey & Company*, 6 December 2022) http://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review> accessed 12 Apr. 2024.

Copyright Act of 1976 and codified in Title 17 of the United States Code.

Copyright Witness Ltd, "A Derivative Work Is a Work That Is Based on (Derived from) Another Work." (*UK Copyright Service*, 11 November 2019) <https://copyrightservice.co.uk/copyright/p22_derivative_works?ssp=1&darkschemeovr=1&s etlang=en&cc=AT&safesearch=moderate> accessed 02 May 2024.

Copyright, Design & Patents Act 1988.

Designers Guild Ltd v Russell Williams (Textiles) Ltd [2000] 1 WLR 2416, [2001] UKHL 17.

EMI Christian Music Grp., Inc. Vs. MP3Tunes, LLC 844 F.3d 79, 89 (2d Cir. 2016).

Fahy L, "Fostering regulator–innovator collaboration at the frontline: A case study of the UK's regulatory sandbox for fintech" (2022) 44 Law & Policy Policy < https://doi.org/10.1111/lapo.12184 > accessed 12 May 2024.

Fisher W, "Reconstructing the Fair Use Doctrine" (1988) 101 HARV.L.REV 1768.

Football Association Premier League Ltd v QC Leisure and Karen Murphy v Media Protection Services Ltd [2012] 1 CMLR 769.

Fox News Network, LLC vs. TV Eyes, Inc. 883 F.3d 169, 178-80 (2d Cir. 2018).

Garner B, *Black's Law Dictionary* (11th edn, Thomson Reuters 2019). *Georgia, et al. vs. Public. Resource. Org, Inc.*, 590 U.S. (2020)

Getty Images (US), Inc. v. Stability AI, Inc., 1:23-cv-00135, (D. Del.)

Getty Images, 'Getty Images Statement' (*Getty Images*, 17 January 2023) <https://newsroom.gettyimages.com/en/getty-images/getty-images-statement>accessed 4 June 2023.

Goyle K, Xie Q, and Goyle V, "DataAssist: A Machine Learning Approach to Data Cleaning and Preparation" (*Arxiv*, 17 July 2023) https://arxiv.org/pdf/2201.07119> accessed 21 Apr. 2024.

Hanger Holdings v Perlake Corporation SA & Anor [2021] EWHC 81 (Ch).

IBM, "What Is Strong AI?" (*IBM*, 13 October 2021) <http://www.ibm.com/topics/strong-ai> accessed 15 Apr. 2024.

IBM, and Morning Consult. "IBM Global AI Adoption Index 2022" (*IBM*, May 2022) https://www.ibm.com/downloads/cas/GVAGA3JP accessed 12 Apr. 2024.

Intellectual Property Office, and Intellectual Property Office "National AI Strategy" (Intellectual Property Office, 2022) https://www.gov.uk/government/publications/national-ai-strategy-html-version> accessed 02 May 2024.

Intellectual Property Office, and UK IPO, "Artificial Intelligence and Intellectual Property: copyright and patents: Government response to consultation" (*Intellectual Property Office*, 2022) https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents-government-response-to-consultation accessed 04 Jun. 2024.

Jaggia S and others, *Business Analytics: Communicating with Numbers* (2nd edn, McGraw Hill 2023).

Jiarui L, "An Empirical Study of Transformative Use in Copyright Law" (2019) 22 STAN. TECH. L. REV. 163 < https://law.stanford.edu/wpcontent/uploads/2019/02/Liu_20190203.pdf > accessed 12 April 2024.

Kadrey v. Meta Platforms, Inc., 3:23-cv-03417, (N.D. Cal.)

Kinicki A and Soignet DB, *Management: A Practical Introduction* (10th edn, McGraw Hill, 2022).

L. v. Alphabet Inc., 3:23-cv-03440, (N.D. Cal. Jul 11, 2023) ECF No. 1.

Liu A. and others "A Survey of Text Watermarking in the Era of Large Language Models" (2023) 1 (1) Arxiv < DOI:<u>10.48550/arXiv.2312.07913</u>> accessed: 14 June 2024.

MAI Systems Corp. v. Peak Computer, Inc., 991 F.2d 511 (9th Cir. 1993).

McKinsey & Company, "What Is Generative AI?" (*McKinsey & Company*, 2 April 2024) http://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai accessed 09 Apr. 2024.

Microsoft Learn "Deep Learning vs. Machine Learning - Azure Machine Learning." (*Deep learning vs. machine learning - Azure Machine Learning / Microsoft Learn*, 19 January 2024) accessed 18 Apr. 2024">https://learn.microsoft.com/en-us/azure/machine-learning/concept-deep-learning-vs-machine-learning?view=azureml-api-2> accessed 18 Apr. 2024.

Montagnon R, "UK Withdraws Plans for Broader Text and Data Mining (TDM) Copyright and Database Right Exception." (*Herbert Smith Freehills*, March 1 March 2023) https://www.herbertsmithfreehills.com/notes/ip/2023-03/uk-withdraws-plans-for-broader-text-and-data-mining-tdm-copyright-and-database-right-exception> accessed 04 Jun. 2024.

Mulder V. and others Trends in Data Protection and Encryption Technologies (Springer 2023)

O'Connor S, "Copyright Law From an American Perspective" (2002) 37 Irish Jurist https://www.jstor.org/stable/44027015 > accessed 12 April 2024.

Oxford Reference, "Limited License," https://www.oxfordreference.com/display/10.1093/oi/authority.20110803100106217 accessed 11 May 2024.

Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146 (9th Cir., 2007).

Public Relations Consultant Association Limited vs. The Newspaper Licensing Agency Limited and others [2013] UKSC 18.

Rao A and others, "Sizing the Prize-What's the Real Value of AI for Your Business and How Can You Capitalize?" (*PWC*, May 2022) <www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>accessed 06 Apr. 2024.

Ringe WG, "Why We Need a Regulatory Sandbox for AI" (Oxford Law Blogs, 12 May 2023) <https://blogs.law.ox.ac.uk/oblb/blog-post/2023/05/why-we-need-regulatory-sandbox-ai> accessed 11 May 2024.

Rudin C, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead" (*Nature Machine Intelligence*, 13 May 2019) https://www.nature.com/articles/s42256-019-0048-x accessed 18 Apr. 2024.

Schroff S, "The Purpose of Copyright-Moving beyond the Theory" (2021) 16 (11) Journal of Intellectual Property Law & Practice https://doi.org/10.1093/jiplp/jpab130> accessed 12 Apr. 2024.

Sega Enterprises Ltd. v. Accolade, Inc., 785 F. Supp. 1392 (N.D. Cal. 1992).

Shavell S, Foundations of Economic Analysis of Law (Harvard University Press 2004).

Silbey J, *Against Progress: Intellectual Property & Fundamental Values in the Internet Age.* (Stanford University Press 2022).

Silverman v. OpenAI, Inc., 3:23-cv-03416, (N.D. Cal.).

Stern Electronics, Inc vs. Kaufman, 669 F.2d 852 (2d Cir. 1982).

Tiffany Design, Inc. v. Reno-Tahoe Specialty Inc., 55 F. Supp. 2d 1113 (D. Nev. 1999).

Tremblay v. OpenAI, Inc., 3:23-cv-03223, (N.D. Cal.).

United States Constitution of 1787.

Wong M and others, "Prompt Evolution for Generative AI: A Classifier-Guided Approach" (*arXiv.org* 24 May 2023) https://arxiv.org/abs/2305.16347> accessed 09 Apr. 2024.

Yuan Z. and others "A blind image watermarking scheme combining spatial domain and frequency domain" (2020) 37 The Visual Computer, 1867–1881 <<u>https://doi.org/10.1007/s00371-020-01945-y</u>.> accessed 14 June 2024.

Zewe A, "Explained: Generative AI" (*MIT News*, 9 November 2003) https://news.mit.edu/2023/explained-generative-ai-1109> accessed 18 Apr. 2024.