32 (2) 2024 IIUMLJ 159 - 192

PROPOSAL FOR COPYRIGHT COMPENSATION FOR ARTIFICIAL INTELLIGENCE (AI) DATA TRAINING IN MALAYSIA

Mohd Syaufiq Abdul Latif*

Nazura Abdul Manap**

Nabeel Mahdi Althabhawi***

ABSTRACT

As Malaysia struggles with challenges presented by the emergence of Artificial Intelligence (AI) in the digital age, there is an increasing need to re-evaluate and potentially revise the country's copyright framework. The Copyright Act 1987 may require adjustments to accommodate the evolving nature of creative works and their production, particularly in the context of AI-generated content, or known as Generative AI. One area of consideration is the implementation of a copyright compensation system which has been successfully adopted in the European Union (EU) and the United States (US) to compensate creators for the use of their works. Hence, this paper explores the feasibility and potential structure of a copyright compensation framework in Malaysia, specifically focusing on compensating rights holders for AI training data used by way of a statutory license and levy system. By examining existing

*PhD Candidate at Faculty of Law Universiti Kebangsaan Malaysia (UKM), 43600, UKM Bangi, Selangor Darul Ehsan, Malaysia. Email: syaufiqlatif@gmail.com (Corresponding Author).

**Professor at Faculty of Law Universiti Kebangsaan Malaysia (UKM), 43600, UKM Bangi, Selangor Darul Ehsan, Malaysia. Email: nazura@ukm.edu.my.

***Assistant Professor at Faculty of Law Universiti Kebangsaan Malaysia (UKM) 43600, UKM Bangi, Selangor Darul Ehsan, Malaysia. Email: althabhawi@ukm.edu.my.

[Received: 17 July 2024, Accepted: 28 October 2024, Published: 31 November 2024]

compensation systems in selected jurisdictions including the EU and the US, this paper aims to provide insights into how such a framework could be effectively implemented into the Copyright Act 1987. The paper also argues that a customised copyright compensation framework could offer a practical solution to the challenges posed by AI, ensuring fair compensation for right holders, promoting innovation, and upholding copyright principles in an increasingly interconnected world. This paper will analyse the current provisions of the Copyright Act 1987, and identify gaps and areas that require reform to effectively address the implications of AI-generated content. This paper finds that the Copyright Act 1987 lacks explicit provisions for compensating rights holders for the use of their works in AI training data, leading to potential gaps in legal protection and fair compensation. As such, the paper recommends specific amendments to the Copyright Act 1987 to incorporate these mechanisms to guide the policymakers in providing a copyright compensation framework to rights holders in Malaysia.

Keywords: Artificial Intelligence, Copyright, Copyright Compensation, Generative AI, Copyright Infringement.

CADANGAN PAMPASAN HAK CIPTA UNTUK LATIHAN DATA KECERDASAN BUATAN (AI) DI MALAYSIA

ABSTRAK

Ketika Malaysia bergelut dengan cabaran yang timbul akibat kemunculan Kecerdasan Buatan (AI) dalam era digital, terdapat keperluan yang semakin mendesak untuk menilai semula dan mengubah kerangka kerja perundangan hak cipta sedia ada di Malaysia. Akta Hak Cipta 1987 di dalam konteks ini memerlukan suatu penyusunan semula dan penyesuaian agar seiring dengan perubahan semasa terutamanya dalam penghasilan apa-apa karya kandungan kreatif yang dihasilkan oleh AI atau yang dikenali sebagai *Generative AI*. Salah satu aspek yang perlu dipertimbangkan dalam penyesuaian ini adalah membentuk suatu sistem kerangka kerja baharu yang dinamakan sebagai 'pampasan hak cipta' di mana sistem ini telah berjaya dibuktikan dan digunapakai di Kesatuan Eropah (EU) dan Amerika Syarikat dalam memberikan suatu akuan pampasan kepada pencipta atas penggunaan karya dalam konteks latihan data AI dalam era digital. Kertas ini secara tidak langsung akan meneroka kebolehlaksanaan sistem ini dalam menstruktur semula potensi sistem rangkakerja 'pampasan hak cipta' di Malaysia dengan memfokuskan khusus pada pemberian pampasan tersebut kepada pemegang hak untuk apa-apa aktiviti 'latihan data AI' yang digunakan

oleh AI melalui pengenalan sistem lesen statutori dan pengenaan levi atas penggunaan tersebut. Kertas ini juga mendapati bahawa Akta Hak Cipta 1987 tidak mempunyai peruntukan jelas dalam memberikan pampasan yang munasabah kepada pemegang hak atas penggunaan karya mereka bagi tujuan 'latihan data AI' yang telah menyebabkan adanya jurang dalam perlindungan undang-undang dan kesesuaian pampasan yang adil kepada pemegang hak cipta. Oleh yang demikian, kertas ini akan mengesyorkan beberapa pindaan kepada Akta Hak Cipta 1987 dengan harapan dapat membantu penggubal dasar dalam menyediakan kerangkakerja pampasan hak cipta yang sesuai dan berkesan kepada pemegang hak cipta di Malaysia.

Kata Kunci: Kecerdasan Buatan, Hak Cipta, Pampasan Hak Cipta, Generative AI, Pelanggaran Hak Cipta.

INTRODUCTION

The integration of AI into various aspects of society has brought about significant changes, particularly in the field of creativity and intellectual property. In Malaysia, as in many other countries, the emergence of AI technology poses unique challenges to the existing copyright framework.[1] Traditional concepts of authorship, ownership, and infringement need to be redefined as AI increasingly contributes to the production of literary, artistic, and musical works.[2] AI's ability to generate original content, mimic human creativity, and autonomously produce works raises fundamental questions about the nature of authorship and the ownership of AI-generated creations. Unlike traditional works, where a human author is identifiable, AI-generated works blur the lines of authorship and ownership leading to uncertainties regarding who should be considered the creator and rightful owner of such works.[3] Furthermore, the widespread use of Generative AI in content creation and dissemination has raised

[1]Marybeth Peters, "The Challenge of Copyright in the Digital Age," *Revista La Propiedad Inmaterial* 9 (2006): 59-67.

[2]Atif Aziz, "Artificial Intelligence Produced Original Work: A New Approach to Copyright Protection and Ownership," *European Journal of Artificial Intelligence & Machine Learning* 2, no. 2 (2023): 3, https://doi.org/10.24018/ejai.2023.2.2.15.

[3]Chawinthorn Watiktinnakorn, Jirawat Seesai, and Chutisant Kerdvibulvech, "Blurring the Lines: How AI Is Redefining Artistic Ownership and Copyright," *Discover Artificial Intelligence* 3 (2023).

concerns about the potential for copyright infringement.[4] AI's ability to process and generate content at scale means that copyrighted works could be reproduced, modified, or distributed without the permission of the rights holders.[5] As a result, the traditional copyright framework, which is designed to protect, incentivise, and compensate human creators, faces significant challenges in adapting to the realities of AI-generated content, particularly for copyrighted works used for the AI data training process.[6]

One area of particular concern is the equitable compensation for creators for the use of their works because of their use in AI data training. The challenges associated with the copyright compensation system for AI data training, including determining the appropriate level of compensation for creators, are important.[7] As AI-generated content becomes more common and widespread, traditional revenue generation methods like royalties and licensing, which have been the mainstay for right holders, may no longer be adequate.[8] This is due to the unique nature of AI-generated works, which can be produced at scale and often without the same level of human effort required for traditional works.[9] The sheer volume and ease of replication of AI-generated content raise questions about how to accurately value these works and ensure that creators receive fair compensation.[10] Thus, the existing revenue

[4]Nicola Lucchi, "ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems," *European Journal of Risk Regulation* (2023): 1–23.

[5]David M. Murray, "Generative AI Art: Copyright Infringement and Fair Use," *SMU Science and Technology Law Review* 26 (2023): 259.

[6]Andres Guadamuz, "Artificial Intelligence and Copyright" (lecture, University of Sussex, United Kingdom, October 2017).

[7]Stichting De Thuiskopie, *International Survey on Private Copying: Law & Practice 2015* (World Intellectual Property Organization, 2016).

[8]Atilla Söğüt, "Dealing with AI-Generated Works: Lessons from the CDPA Section 9(3)," *Journal of Intellectual Property Law & Practice* 19, no. 1 (2024): 43–54.

[9]Jenny Quang, "Does Training AI Violate Copyright Law?" *Berkeley Technology Law Journal* 36, no. 4 (2021): 1407

[10]Paul Keller, "Protecting Creatives or Impeding Progress? Machine Learning and the EU Copyright Framework," February 17, 2023, https://copyrightblog.kluweriplaw.com/2023/02/20/protecting-

streams may not suffice to compensate creators for the use of their works in AI data training, highlighting the need for new approaches to address these challenges.

At the beginning, this article would be incomplete without first addressing the issue of creators not being compensated for the use of their work by AI companies. In Malaysia, several issues have been reported indicating that the use of AI in creating works raises questions about whether rights holders in Malaysia should be compensated and how this compensation should be determined.[11] As AI generates images, including its use in advertising, product photos, and AI art, which can provide new sources of income for businesses and entrepreneurs in Malaysia, there are concerns regarding the issue of ownership of images and videos created by AI where the laws need to be established to respect artists and their works.[12] It is also evident that the law is lagging in effectively addressing the challenges faced by creators in the digital age.[13] The Copyright Act 1987 in Malaysia, for example, does not specifically address the use of AI in creating works, leading to uncertainties regarding the rights of creators and the obligations of AI companies to compensate them for the use of their work. The copyright infringement on the other hand is a serious concern that must be addressed to protect the rights of creators.[14]AI's ability to replicate and distribute copyrighted works at scale raises questions about how to effectively enforce copyright laws in the digital

creatives-or-impeding-progress-machine-learning-and-the-eu-copyright-framework/.

[11]FMT Reporters. (2024, February 22). Set up task force to deal with AI use, govt told. Free Malaysia Today. Retrieved from https://www.freemalaysiatoday.com/category/nation/2024/02/22/set-up-task-force-to-deal-with-ai-use-govt-told/, accessed April 9, 2024.

[12]Marlia Zakaria, "MyIPO Bakal Bangkitkan Lagi Isu Penggunaan 'Copyright' AI di Geneva," *Kosmo*, February 15, 2024, https://www.kosmo.com.my/2024/02/15/myipo-bakal-bangkitkan-lagi-isu-penggunaan-copyright-ai-di-geneva/, accessed April 9, 2024.

[13]Nurqalby Mohd Reda, "Malaysia Juga Berdepan Risiko Jenayah AI - Pakar," *Bernama*, February 5, 2024, https://bernama.com/bm/rencana/news.php?id=2268227.

[14]Utusan Malaysia. (2023, Disember 22). Bimbang AI tidak hormati budaya Malaysia. Utusan Malaysia. Retrieved from https://www.utusan.com.my/rencana/forum/2023/12/bimbang-ai-tidak-hormati-budaya-malaysia/, accessed May 15, 2024.

age and the need to amend existing laws related to technology and crime to keep pace with current developments to remedy the creators in any way possible.[15]

Considering these issues and challenges, it is essential to develop a comprehensive copyright compensation framework in Malaysia that specifically addresses the use of copyrighted works particularly in compensating right holders in AI data training. It is to be noted that the Copyright Act 1987 in Malaysia does not provide any provision allowing the use of copyrighted works in AI data training or compensation systems for right holders, leading to uncertainties regarding the rights of creators and the obligations of AI companies. This lack of clarity in current provisions can create challenges for creators seeking fair compensation for the use of their works in AI data training. Therefore, this paper will focus on exploring the feasibility and potential structure of a copyright compensation framework for Malaysia in the age of AI, with a particular emphasis on the use of copyrighted works for AI data training.

The paper begins with an analysis of the current copyright landscape in Malaysia, followed by a discussion on the unauthorised use of copyrighted works to train AI systems. Next, it examines the right to compensation for creators whose works are used in AI data training. The paper then provides an overview of the existing copyright compensation systems before proposing a copyright compensation framework tailored for Malaysia. Key sections also include a discussion of eligibility criteria for compensation and the feasibility of implementing a hybrid compensation framework under the Copyright Act 1987. The paper concludes with recommendations on how to adapt the existing legal framework to ensure fair compensation for creators in the AI age.

[15]The Star. (2024, February 15). AI-produced content resembling works registered under intellectual property a violation, says MyIPO. Retrieved from https://www.thestar.com.my/news/nation/2024/02/15/ai-produced-content-resembling-works-registered-under-intellectual-properly-a-violation-says-myipo, accessed June 3, 2024.

RESEARCH METHODOLOGY

This paper adopts a black letter law approach, focusing on legal doctrinal and comparative law techniques through library research. The research includes a review of statutes, extra-legal materials, books, articles, newspapers, and seminar papers. The analysis used in this research includes qualitative methods of textual and thematic analysis to examine legal texts, categorise and interpret provisions, and analyse themes within the jurisdictions of Malaysia, the EU, and the US. This methodological approach is chosen to inform potential reforms to the Copyright Act 1987 in Malaysia, ensuring a comprehensive understanding of copyright compensation in the context of AI training data. It also aims to analyse the legal framework for compensating rights holders for the use of their copyrighted works as AI training data in Malaysia, with a focus on the Copyright Act 1987, while comparing it to the frameworks in the EU and US.

In this paper, the researchers will also compare both substantive and procedural law aspects. Substantive law analysis focuses on copyright compensation provisions, while procedural law analysis examines the implementation mechanisms of both statutory licenses and levy systems. The criteria for comparison include similarities, differences, and unique features currently in place from the perspectives of the EU and the US. The EU and the US are specifically chosen for their advanced legal frameworks in copyright protection, as both jurisdictions provide a basis for this proposal, which will be discussed in detail in the later part of this paper.

CURRENT COPYRIGHT LANDSCAPE

The current state of AI integration in Malaysia is set apart by a growing recognition of its potential to drive innovation and economic growth across various sectors, including creativity and intellectual property.[16]

[16]Angeline Yeoh and Christoper Fam, "Moving Forward with AI: How Malaysia Can Grow into an AI-Savvy Nation with the Right Talent and Innovation," *The Star*, August 14, 2023, https://www.thestar.com.my/tech/tech-news/2023/08/14/moving-

Malaysia has made strides in embracing AI technologies, with initiatives such as the National AI Framework and the National Policy on Industry 4.0 highlighting the government's commitment to fostering AI development. In the field of creativity and intellectual property, AI is increasingly being used to enhance the production and dissemination of copyrighted works.[17] AI technologies such as machine learning and natural language processing have enabled the creation of AI-generated content, including music, literature, and visual art.[18] These developments raise important questions about the role of AI in the creation process and the rights of creators in AI-generated works.

The current copyright landscape in Malaysia, as governed by the Copyright Act 1987, does not specifically address the challenges posed by AI-generated content, particularly in the context of AI data training. AI data training involves using copyrighted works as input to train AI models, which can then generate new works that may be considered derivative works or even original works.[19] The Copyright Act 1987 also does not clearly define the rights and obligations related to AI data training, leaving a legal grey area regarding the use of copyrighted works in this context. This lack of clarification raises concerns about the potential infringement of copyright holders' rights and the need for mechanisms to ensure fair compensation for the use of copyrighted works in AI data training in Malaysia. Without clear guidelines, creators may be hesitant to contribute their works to AI training data sets, which could hinder the development of AI technologies and innovation in Malaysia.[20]

---

forward-with-ai-how-malaysia-can-grow-into-an-ai-savvy-nation-with-the-right-talent-and-innovation, accessed June 18, 2024.

[17]Guadamuz Andres, "Artificial Intelligence and Copyright" (lecture, University of Sussex, United Kingdom, October 2017).

[18]Zhihan Lv, "Generative Artificial Intelligence in the Metaverse Era," *Cognitive Robotics* 3 (2023): 208–217, http://www.keaipublishing.com/en/journals/cognitive-robotics/, accessed March 29, 2024.

[19]Quang, "Training AI," 1407.

[20]Joan S. Santani. Santani, "Malaysia Coming Up with AI Governance and Code of Ethics," *New Straits Times*, January 9, 2024, https://www.nst.com.my/business/corporate/2024/01/998905/malaysia-coming-ai-governance-and-code-ethics, accessed March 31, 2024.

Historically, Malaysia also has not implemented a copyright compensation system, unlike some other countries that have adopted such systems to compensate rights holders for the use of their works in various ways, such as through copyright levies or statutory licenses.[21] The absence of a copyright compensation system in Malaysia means that rights holders may not receive fair compensation for the use of their works in the digital environment and also to AI data training process, potentially discouraging them from contributing their works to AI projects. This could hold back innovation and creativity in Malaysia's AI sector, as creators may be less inclined to share their works for fear of not being adequately compensated. Without appropriate laws to address the compensation for the use of their works in AI data training, creators may face challenges in protecting their rights and receiving fair compensation for their contributions.[22]

USING COPYRIGHTED WORK TO TRAIN AI DATA WITHOUT PERMISSION

Training AI models using copyrighted works involves using these works as input data to teach the model to perform certain tasks or generate new content.[23] This process is essential for developing AI systems that can understand, interpret, and generate human-like responses or creative works.[24] However, the use of copyrighted works for AI training raises several legal and ethical considerations. One of the main concerns is whether such use constitutes copyright infringement.[25] Copyright law generally grants creators the exclusive

[21]Stichting De Thuiskopie, *International Survey.*

[22]Martin Senftleben, (2023). Generative AI and Author Remuneration. *IIC - International Review of Intellectual Property and Competition Law*, 54(4), 1535–1560. https://doi.org/10.1007/s40319-023-01399-4

[23]Christopher T. Zirpoli, "Generative Artificial Intelligence and Copyright Law," *Copyright, Fair Use, Scholarly Communication* (2023): 1-5.

[24]Zhuohao Wu, Danwen Ji, Kaiwen Yu and Mohammad Shidujaman, "AI Creativity and the Human-AI Co-creation Model," in *Human-Computer Interaction: Theory, Methods and Tools. HCII 2021. Lecture Notes in Computer Science*, vol. 12762, ed. M. Kurosu (Springer, Cham, 2021).

[25]Andrew W. Torrance and Beth Tomlinson, "Training Is Everything: Artificial Intelligence, Copyright, and Fair Training," *Dickinson Law Review*, forthcoming 2023.

right to reproduce, distribute, and display their works.[26] When copyrighted works are used to train AI models, they are often reproduced and transformed in the process, potentially implicating these exclusive rights.[27] Malaysia and countries with obligation under the World Intellectual Property Organisation (WIPO) Treaties, and the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), copyright law allows the member states to incorporate in their domestic law exceptions or limitations to copyright infringement for certain purposes, such as research, education, or criticism.[28] Whether using copyrighted works for AI training falls under these exceptions depends on the specific circumstances and the laws of the member states.

Another consideration is the impact of using copyrighted works for AI training on the rights and interests of the copyright holders.[29] While the use of copyrighted works for AI training may not directly compete with the market for the original works, it could potentially affect the value of the works or the ability of creators to control their works' use.[30] To address these concerns, the authors will examine the selected EU Directive and US Copyright Act 1976 which have existing mechanisms in place to extend the existing copyright compensation. In the EU levies exist and statutory licenses established in the US to compensate rights holders. There are proposals for the extension of the use of their works in AI training.

While the Copyright Act 1987 does not incorporate any provision related to AI data training specifically, the Act however does provide for certain exceptions to copyright infringement that may apply in this context. For example, Section 13(2)(a) of the Act allows for the

[26] Khaw Lek Tee, "Copyright Law in Malaysia: Does the Balance Hold?" *Journal of Malaysian and Comparative Law* (2004).

[27]Martin Kretschmer, Tzania Margoni, and Pinar Oruç, "Copyright Law and the Lifecycle of Machine Learning Models," *IIC - International Review of Intellectual Property and Competition Law* 55 (2024): 110–138

[29]Congressional Research Service, "Generative Artificial Intelligence and Copyright Law," Congressional Research Service, September 29, 2023, https://crsreports.congress.gov/product/pdf/LSB/LSB10922, accessed February 21, 2024.

[30]Quang, "Training AI," 1407.

reproduction of copyrighted works for the purpose of research or private study. This exception could potentially cover the use of copyrighted works for AI training if the use is considered to be for research purposes. Additionally, Section 13(2)(b) of the Act allows for the reproduction of copyrighted works for the purpose of criticism or review. This exception could apply if the use of copyrighted works for AI training is considered to be a form of criticism or review of the works themselves. However, the application of these exceptions to AI training is not straightforward and would depend on the specific facts of each case in Malaysia. As such, it is suggested that the complexities inherent in these cases may necessitate introducing new exceptions or limitations to copyright infringement specifically designed for AI training or providing additional guidance on how existing exceptions apply in this context.

THE RIGHT TO BE COMPENSATED AS A RESULT OF AI DATA TRAINING

The right to be compensated as a result of AI data training refers to the concept that creators and rights holders should receive fair compensation when their copyrighted works are used to train AI models.[31] AI models are algorithms or systems that are trained on large datasets, including copyrighted works, to learn patterns and generate new content.[32] These models rely on the underlying data to improve their performance and create outputs, such as text, images, or music.[33] As AI technology becomes more prominent in various industries, including the creative sector, the use of copyrighted works for AI training has raised concerns about the impact on creators' rights and the need for compensation.[34] Governments around the world have also faced massive procedural ethical, and economic challenges in

[31]Senftleben, "Generative AI," 1535–1560

[32]Kretschmer et al., "Copyright Law," 110–138.

[33]Andrew W. Torrance and Beth Tomlinson, "Training Is Everything: Artificial Intelligence, Copyright, and Fair Training," *Dickinson Law Review*, forthcoming 2023.

[34]Nicola Lucchi, "ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems," *European Journal of Risk Regulation* (2023): 1–23.

determining how to fairly compensate rights holders for AI data training.[35] One of the key challenges is defining the scope of the right to be compensated. Should all use of copyrighted works for AI training be subject to compensation or should there be exceptions for certain types of uses?[36]

In addition, determining the appropriate rate of compensation presents a significant challenge. The rate must be sufficient to provide fair compensation to rights holders, but it must also be reasonable enough not to hinder innovation in AI development.[37] Another challenge is identifying the mechanisms for collecting and distributing compensation. Copyright levies have been proposed as a possible solution, but implementing such a system requires careful consideration of legal, economic, and practical factors.[38] As the problem and issues relating to compensation to creators worsen, governments and stakeholders must work together to address the challenges and develop effective solutions. One approach is to establish a comprehensive legal framework that clearly defines the rights and obligations of rights holders, AI developers, and users in the Copyright Act 1987 itself. Such a framework should include provisions for fair compensation of rights holders, mechanisms for collecting and distributing compensation, and enforcement measures to ensure compliance.[39]

---

[35]Blake Brittain, "Getty Images Lawsuit Says Stability AI Misused Photos to Train AI," *Reuters*, February 6, 2023, https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06/, accessed May 29, 2024.

[36]Maximiliano Marzetti, "Incorporating Exceptions and Limitations to Copyright Law to Incentivize the Development of Artificial Intelligence in Latin America," *Latin American Law Review* (2022): 19-33.

[37]Christophe Geiger and Vincenzo Iaia, "The Forgotten Creator: Towards a Statutory Remuneration Right for Machine Learning of Generative AI," *Computer Law & Security Review* 52 (2024), https://doi.org/.

[38]Christinaristina Angelopoulos and João Pedro Quintais, "Fixing Copyright Reform: A Better Solution to Online Infringement," *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law* 10, no. 2 (2019).

[39]Philip Hacker, "A Legal Framework for AI Training Data—From First Principles to the Artificial Intelligence Act," *Law, Innovation and Technology* 13, no. 2 (2021): 257-301.

This article would be lacking if it did not include some of the legal uncertainties caused by AI data training on copyright works. For example, on 27 December 2023, the New York Times (a newspaper company) filed a lawsuit against OpenAI and Microsoft for copyright infringement, alleging that millions of its articles were used to train AI chatbots that now compete with the newspaper for providing reliable information.[40] The lawsuit, filed in Federal District Court in Manhattan, seeks billions of dollars in damages and calls for the destruction of chatbot models using copyrighted material from the New York Times. Another case was also reported on October 23, 2023, where Universal Music filed a lawsuit against Anthropic, the AI startup behind Claude the chatbot, for alleged copyright infringement of song lyrics.[41] The lawsuit, filed in a Tennessee federal court, accuses Anthropic of "systematic and widespread infringement" of copyrighted lyrics by generating "almost identical copies" of songs such as "Roar" by Katy Perry and "I Will Survive" by Gloria Gaynor. Another case also occurred recently on March 12, 2024, where tech giant Nvidia faced a lawsuit from authors Brian Keene, Abdi Nazemian, and Stewart O'Nan, who alleged that Nvidia used their 196,640 copyrighted books without its permission to train its AI platform NeMo.[42] In the United Kingdom, on 7 February 2023, Getty Images filed a lawsuit against Stability AI in the High Court of Justice in London, alleging infringement of intellectual property rights, including copyright in Getty Images' content.[43] Getty Images claims that Stability AI unlawfully copied and processed millions of images and associated metadata without a license, benefiting Stability AI's commercial interests and harming content creators.

[40]*The New York Times*, "Microsoft Seeks to Dismiss Parts of Suit Filed by The New York Times," March 4, 2024, https://www.nytimes.com/2024/03/04/technology/microsoft-ai-copyright-lawsuit.html.

[41]Hayden Field, "Anthropic's AI Chatbot Claude Is Posting Lyrics to Popular Songs, Lawsuit Claims," *CNBC*, October 18, 2023, https://www.cnbc.com/2023/10/18/universal-music-sues-anthropic-for-copyright-infringement-over-lyrics.html, accessed June 17, 2024.

[42]Jonathan Stempel, "Nvidia Is Sued by Authors Over AI Use of Copyrighted Works," *Reuters*, March 10, 2024, https://www.reuters.com/technology/nvidia-is-sued-by-authors-over-ai-use-copyrighted-works-2024-03-10/, accessed June 217, 2024.

[43]Brittain, "Getty Images Lawsuit."

The lawsuits filed by the right holders signify a growing trend of legal challenges related to legal uncertainties surrounding the use of copyrighted materials in training AI models, especially in cases where the AI models generate content that closely resembles the original works. The lawsuits indicate a need for clearer legal and regulatory frameworks governing the use of copyrighted works in AI development particularly in the Malaysian context. The establishment of such frameworks could help prevent future disputes and provide guidance on how to direct the complexities of AI and data rights through the copyright law framework. As such, the authors in this paper suggest a copyright compensation system as a potential solution to compensate rights holders for the use of their works in AI training particularly in the Copyright Act 1987 context.

OVERVIEW OF COPYRIGHT COMPENSATION SYSTEMS

Shaping a copyright compensation system as proposed in this paper, could provide a mechanism for compensating rights holders for the use of their works in AI data training. This approach is particularly relevant in the context of AI development due to its far-reaching use not only in Malaysia but the globe at large. By establishing a clear framework for compensating rights holders, Malaysia can ensure that creators are fairly remunerated for the use of their works while also promoting innovation and creativity in the AI sector. The following discussion will thus provide an overview of the crucial elements that may shape the possible framework of the copyright compensation system in Malaysia.

EU Copyright Levy Approach

In the EU, copyright levy systems are employed to compensate rights holders for the use of their works in various contexts not covered by traditional copyright licensing schemes.[44] The EU Copyright and Information Society Directive 2001/29 (Infosoc Directive) recognises the need for fair compensation for rights holders in cases of exceptions or limitations to copyright. For example, Article 5(2)(b) of the EU Infosoc Directive allows member states to introduce an exception or limitation to the reproduction right for private copying, provided that

[44]Stichting De Thuiskopie, *International Survey.*

rights holders receive fair compensation. To ensure this fair compensation, many EU member states have implemented copyright levy systems.[45] These systems impose levies on devices or services capable of reproducing copyrighted works, such as blank media, recording devices, or digital storage media.

Some EU countries such as Germany, France and Italy have more extensive levy systems that cover a wide range of devices and media, while others such as Netherlands, Denmark and Ireland have more limited systems that focus on specific types of devices or media.[46] Generally, the idea is that the imposition of copyright levy in the EU revolves around "private use" copying.[47] In the EU context, private use refers to the act of individuals making copies of copyrighted works for their personal, non-commercial use.[48] This means that individuals are permitted to make copies of copyrighted works for their private use, such as making a backup copy of a CD or DVD, without infringing copyright law.[49] When making copies for private use, individuals are expected to pay a levy on the devices or media used for reproduction, the revenue from these levies is then distributed to rights holders through collecting societies that represent the right holders.[50] This levy is to ensure that rights holders are fairly compensated for the use of their works in private copying situations, while also allowing

[45]JoostPoort, "Copyright Levies," in *Handbook of the Digital Creative Economy*, eds. R. Towse and C. Handke (Edward Elgar, 2013), 235-246.

[46]Stichting De Thuiskopie, *International Survey.*

[47]João Pedro Quintais, "Private Copying and Downloading from Unlawful Sources," *International Review of Intellectual Property and Competition Law* (IIC), Institute for Information Law Research Paper No. 2014-03, Amsterdam Law School Research Paper No. 2014-55, 2015.

[48]Christoph Peukert, "Copyright Levies and Cloud Storage: Ex-Ante Policy Evaluation with a Field Experiment," *Research Policy* 53 (2024): 104918.

[49]Martin Kretschmer, "Private Copying and Fair Compensation: An Empirical Study of Copyright Levies in Europe," Intellectual Property Office Research Paper No. 2011/9, University of Glasgow - School of Law, 2011.

[50]Poort, "Copyright Levies," 235–246

individuals to enjoy reasonable access to copyrighted content for personal use.[51]

It is to be noted that the purpose of copyright levies is to provide rights holders with a source of income for the use of their works in situations where it may be difficult or impractical to obtain individual licenses.[52] In such cases, copyright levies provide a practical solution by collecting fees from the sale of devices and media that can be used for private copying, and distributing these fees to rights holders thus providing rights holders with a means of revenue to compensate for the use of their work by third parties and maintain a sustainable ecosystem where right holders are incentivised to continue producing new works.[53]

US Statutory Licensing Approach

In contrast to the approach taken by some EU countries, the US has opted for a more targeted approach to compensating rights holders for the use of copyrighted works. Rather than implementing a broad copyright levy system that imposes fees on devices or media capable of reproducing copyrighted content, the US has developed specific statutory licenses for the use of copyrighted works.[54] One example of this approach is the statutory license provided under section 17 U.S.C. § 114 of the US Copyright Act. Section 17 U.S.C. § 114 of the US Copyright Act allows certain types of digital music services, such as internet radio stations, to publicly perform sound recordings without

[51]Jane C. Ginsburg, "Fair Use for Free, or Permitted-but-Paid?" *Berkeley Technology Law Journal* 29, no. 3 (2015): 1383-1446.

[52]Stanley M, Besen., Sheila. N. Kirby, and Steven, "An Economic Analysis of Copyright Collectives," *Virginia Law Review* 78, no. 1 (1992): 383-411.

[53]Reto M. Hilty and Valentina Moscon, eds., *Modernisation of the EU Copyright Rules: Position Statement of the Max Planck Institute for Innovation and Competition*, Max Planck Institute for Innovation and Competition Research Paper No. 17-12 (Munich: Max Planck Institute for Innovation and Competition, 2017), https://doi.org/10.17617/2.2470998.

[54]Jacob Noti-Victor, "Reconceptualizing Compulsory Copyright Licenses," *Stanford Law Review* 72 (2020): 915.

having to negotiate directly with each rights holder.[55] Another example of a statutory license in the US Copyright Act is the license for making and distributing phonorecords of nondramatic musical works under section 17 U.S.C. § 115. This provision allows anyone to obtain a license to make and distribute phonorecords of a musical work without the permission of the copyright owner.[56]

Similarly, the US Copyright Act includes a compulsory license under section 17 U.S.C. § 115 for the mechanical reproduction and distribution of musical compositions. This license allows anyone to cover a song without having to negotiate directly with the copyright owner. These statutory licenses offer a streamlined approach for obtaining permission to use copyrighted works in specific ways or specific procedures, such as serving notice and paying royalties at statutory rates set by the Copyright Royalty Board.[57] While the US approach does not involve a broad copyright levy system like those found in some EU countries, it achieves a similar goal of ensuring that rights holders receive compensation for the use of their works in specific situations. This targeted approach allows for more flexibility and control over the licensing process, while still providing a mechanism for rights holders to be fairly compensated for the use of their works.

The Debates in the US and EU

While there have been extensive discussions and proposals for implementing a copyright compensation system through the EU copyright levy and the US statutory license to extend the application to AI data training, these proposals have not been widely adopted and

[55]Howard. B. Abrams, "Copyright's First Compulsory License," *Santa Clara High Technology Law Journal* 26 (2009): 215.

[56]Yolanda King, "The Inadvisability of Nonuniformity in the Licensing of Cover Songs," *Belmont Law Review* 3 (2016): 51.

[57]Brian Sanchez, "The Section 115 Mechanical License and the Copyright Modernization Act: The Hardships of Legislating Music Industry Negotiations," *DePaul Journal of Art, Technology & Intellectual Property Law* 17 (2006): 37

remain a suggestion and idea.[58] Critics argue that such levies or statutory license suggestions could stifle innovation and impose additional costs on consumers without providing clear benefits to rights holders. The situation is similar in many jurisdictions where there is an ongoing debate about the feasibility and effectiveness of implementing copyright levies or statutory licenses for AI training data.[59] Some argue that such levies could help ensure that rights holders are fairly compensated for the use of their works in AI development.[60] However, others raise concerns about the potential impact on innovation and the complexity of implementing and enforcing such a system. While copyright levies or statutory licenses have been successful in some contexts, such as private copying of music and blank media, extending this compensation to AI training data poses unique challenges because the legal and regulatory framework surrounding AI and copyright is still evolving.[61]

At the outset, it is important to note that Malaysia does not have a similar provision regarding copyright levy either in the form of a broad specific levy system in the EU or specific statutory license on any devices or services that can reproduce copyrighted works, including those used for AI training in the Copyright Act 1987. It is indeed plausible to consider the introduction of a copyright levy system in place as AI technology continues to advance and its impact on copyright becomes more pronounced. Implementing such a copyright levy could help ensure that rights holders are fairly compensated for the use of their works and in particular a copyrighted work used to train AI data while also promoting innovation and creativity in the AI sector. This is to note that, any decision possibly shaping copyright levy in Malaysia would need to be carefully considered in the context of Malaysia's legal and regulatory framework, considering the interests of

[58]Christophe Geiger and Valeria, "The Forgotten Creator: Towards a Statutory Remuneration Right for Machine Learning of Generative AI," *Computer Law & Security Review* 52 (2024)

[59]Kretschmer et al., "Copyright Law," 110–138.

[60]Senftleben, "Generative AI," 1535–1560

[61]Gerrit De Vynck, "AI Learned from Their Work. Now They Want Compensation," *The Washington Post*, July 16, 2023, https://www.washingtonpost.com/technology/2023/07/16/ai-programs-training-lawsuits-fair-use/, accessed June 2, 2024.

rights holders, AI developers, and consumers that will be discussed below.

PROPOSED COPYRIGHT COMPENSATION FRAMEWORK FOR MALAYSIA

To address the challenges posed by AI in the digital environment, particularly regarding the compensation of creators for the use of their works by AI companies, the copyright compensation framework could be implemented within the existing legal and regulatory framework for copyright in Malaysia. This would include amendments to the Copyright Act 1987 to include provisions for the compensation of rights holders for the use of their works in AI data training. One approach that Malaysia could consider is a system similar to the EU copyright levy approach, where levies are imposed on devices or services capable of reproducing copyrighted works. This would require amendments to Part III of the Copyright Act 1987 to include provisions for such a levy system. Part III of the Copyright Act 1987, for instance, includes provisions on the rights of copyright owners, the scope of these rights, and the exceptions to copyright infringement. Amendments could include expanding the scope of copyright to specifically address issues related to AI data training. Alternatively, Malaysia could adopt a statutory licensing approach akin to the US, where specific licenses are provided for the use of copyrighted works in certain circumstances. This would involve amending section 13(2) of the Copyright Act 1987 to include provisions for statutory licenses for the use of copyrighted works in AI data training. In this paper, the authors suggest a hybrid approach where Malaysia could implement a system that includes elements of both the EU copyright levy approach and the US statutory licensing approach. Such key components for the hybrid system would be as follows:

Definition, Scope, and Type of Works of AI Data Training

In considering the scope of a potential copyright compensation framework within the Copyright Act 1987, several key aspects need to be addressed. One crucial area is the definition of AI data training and

the types of uses that would be covered under the framework.[62] This could include the scope of use of copyrighted works for training machine learning algorithms, developing AI models, and other similar purposes.[63] Another important aspect to consider is the types of work that would be covered under the framework. This could include literary, musical, and artistic works, as well as films and sound recordings, which are often used in AI data training. For example, the current section 7 of the Copyright Act 1987 in Malaysia defines literary works as including computer programmes, compilations of data, and databases, which are relevant to AI data training. However, the Act does not specifically address the use of copyrighted works for AI data training purposes. Therefore, there is a need to clarify the scope of the framework to ensure that it covers all relevant uses of copyrighted works in AI data training.

Right Holders' rights to AI data training

The framework should clarify the rights of rights holders concerning AI data training, including the exclusive rights to control reproduction, communication to the public, performance, distribution, and commercial rental of copyrighted works.[64] Similarly, the framework should address the issue of fair compensation for rights holders. The statutory license could include provisions for the calculation and payment of royalties, as well as mechanisms for monitoring and enforcing compliance with the license terms.[65] This is to ensure that rights holders are adequately compensated for the use of their works in AI data training, while also providing a clear framework for users to obtain the necessary licenses.

[62]Hacker, "Legal Framework," 257–301

[63]Theodoros Chiou, (2019). Copyright lessons on Machine Learning: what impact on algorithmic art? *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law*, 10(3).

[64]Haochen Sun, "Redesigning Copyright Protection in the Era of Artificial Intelligence," *Iowa Law Review* 107 (2022): 1213.

[65]David. W. Opderbeck, "Copyright in AI Training Data: A Human-Centered Approach," *Oklahoma Law Review* 76 (2024).

Introduction of statutory license provision

Apart from the definition, scope, types of works, and introduction of right holders' rights to be incorporated in the Copyright Act 1987, Malaysia could introduce a new statutory licensing scheme for the use of copyrighted works in AI data training. This would complement the existing framework and provide a more efficient mechanism for obtaining licenses for such uses. The statutory licensing scheme could be modelled after existing provisions in other jurisdictions, such as the US statutory licenses for music streaming services, which provide a clear and standardised process for obtaining licenses and paying royalties.[66] This approach would be similar in principle to the EU levy system, where levies are imposed on devices or services capable of reproducing copyrighted works. However, instead of imposing levies on specific devices or services, the statutory licensing scheme would establish a framework for AI companies to obtain licenses for the use of copyrighted works in AI data training. In turn, the collection in the form of levies would be collected and distributed to right holders through a designated Collective Management Organisation in Malaysia.

ELIGIBILITY CRITERIA

The eligibility criteria for obtaining a license to use copyrighted works in AI data training under the proposed hybrid levy and statutory licensing scheme as discussed above would be a crucial aspect of the framework. The process would need to be clear, transparent, and accessible to ensure that AI companies can obtain the necessary licenses efficiently while also protecting the rights holders' rights.[67] Key components must be considered to ensure the effectiveness and fairness of the compensation scheme. The eligibility criteria for obtaining a license hence could include transparency on payment

[66]Guadamuz Andres, "Artificial Intelligence and Copyright," lecture, University of Sussex, October 2017.

[67]Daniel Rodriguez Maffioli, "Copyright in Generative AI Training: Balancing Fair Use through Standardization and Transparency," August 21, 2023, SSRN, https://ssrn.com/abstract=4579322, accessed 19 June, 2024.

collection and distribution of royalties collected under the hybrid scheme.

Payment of Royalties

AI companies would be required to pay royalties in the form of levies if a statutory license is implemented for the use of copyrighted works in AI data training. The levies would be collected from AI companies based on their usage of copyrighted works for training AI models and algorithms.[68] The amount of the levies could be determined based on factors such as the type and quantity of copyrighted works used, the commercial value of the AI applications developed using the works, and the potential impact on the market for the original works as enshrined under section 13(2A) of the Copyright Act 1987. The specific details of how such a levy would be calculated and implemented would need to be further developed and specified in the legislative framework, considering the principles of fairness, transparency, and effectiveness.[69] In cases where the works used are not quantifiable, such as when the works are used in a transformative or derivative manner that makes it difficult to determine the specific copyrighted content used, a flat-rate levy or a percentage-based levy could be applied.[70] This approach would provide a simpler and more practical method for calculating royalties "per use". It would also ensure that rights holders receive fair compensation for the use of their works in AI data training, even in cases where the specific works used are not easily quantifiable.

Collection and Distribution

In Malaysia, there are existing Collective Management Organisations (CMOs) that have been declared under Section 27A of the Copyright Act 1987. These CMOs are responsible for licensing rights in

[68]Martin Senftleben, "AI Act and Author Remuneration - A Model for Other Regions? Institute for Information Law (IViR), University of Amsterdam, 2024.

[69]Quang, "Training AI," 1407.

[70]Nicola Lucchi, "ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems," *European Journal of Risk Regulation* (2023): 1–23.

copyrighted works and collecting royalties on behalf of rights holders. If Malaysia was to implement a copyright compensation system for AI data training, these existing CMOs could potentially be authorised to collect the levies, given their established roles in managing copyright licensing and royalties. However, amendments to the Copyright Act 1987 would likely be needed to expand the scope of these CMOs' authority under its existing declaration to include the collection of levies for AI data training. Additionally, to facilitate the collection process, the CMOs could establish agreements with manufacturers, importers, and distributors of devices and services used for AI data training to collect the levies at the point of sale or distribution.[71] This could be done through a licensing scheme or a similar mechanism that allows these entities to use copyrighted works in AI training in exchange for paying the levies. The organisation could also establish a reporting system that requires entities using copyrighted works for AI data training to report their usage and pay the levies accordingly. This is to ensure that all relevant entities are contributing to the compensation of rights holders and provide transparency in the collection process. The CMOs would also need to work closely with AI developers and users to raise awareness about the levy system and the importance of compensating rights holders for the use of their works. This could include preparing or establishing guiding compliance and best practices for using copyrighted works in AI data training.[72]Transparency in the distribution process is also another important area to be checked. Rights holders should have access to information about how their compensation is calculated and the eligibility criteria used for distribution in the payment of royalties.[73] This transparency in collection and distribution thus helps build trust and ensures that rights holders are fairly compensated for their contributions, particularly in AI data training.

[71]CISAC, "Remuneration, Authorisation, and Transparency: CISAC Outlines Priorities for Creators in AI Age," , https://chat.openai.com/c/aa13436a-43b8-42f9-b19e-223d8b7d7287, accessed 7 April 2024.

[72]Andrew W. Torrance and Beth Tomlinson, "Training is Everything: Artificial Intelligence, Copyright, and Fair Training," *Dickinson Law Review*, forthcoming 2023.

[73]Ginsburg, "Fair Use," 1383–1446.

FEASIBILITY OF THE PROPOSED HYBRID COMPENSATION FRAMEWORK UNDER COPYRIGHT ACT 1987

Based on the previous discussion, the proposed compensation framework for AI data training, including the introduction of a statutory licensing scheme and potential copyright levy system, could be workable under the Copyright Act 1987 of Malaysia. The Berne Convention for the Protection of Literary and Artistic Works ("Berne Convention") and Trade-Related Aspects of Intellectual Property Rights Agreement ("TRIPS") which Malaysia is a party, allows member countries to establish limitations and exceptions to copyright (i) in certain special cases as long as (ii) they do not conflict with the normal exploitation of the work and (iii) do not unreasonably prejudice the legitimate interests of the author (the three steps test). The proposed framework, which aims to facilitate AI data training while ensuring fair compensation for rights holders is in line with these principles. The TRIPS Agreement also requires member countries to provide adequate and effective protection of intellectual property rights, including copyright. In the context of AI training data, a copyright compensation framework could be seen as an exception or limitation to copyright if it is designed to facilitate access to copyrighted works to train AI models while respecting right holders' exclusive rights.[74] However, such a framework would need to comply with the three-step test. For example, a copyright compensation framework for AI training data could be designed to apply only to certain types of copyrighted works or to certain uses of those works as proposed in this paper.

In the EU for instance, the use of copyrighted works for AI data training is subject to the EU Copyright Directive 2019/790/EU on Copyright and Related Rights In The Digital Single Market ("DSM Directive"). Under the DSM Directive, Member States have the option to implement exceptions or limitations to copyright for text and data

[74]Erin Keenan-O'Malley, "'AI' and 'Copyright Infringement' Increasingly Feature in the Headlines, but What Are the Legal Issues Being Debated?" *EIP Newsletter*, UK, December 15, 2023,https://www.aippi.org/news/ai-and-copyright-infringement-increasingly-feature-in-the-headlines-but-what-are-the-legal-issues-being-debated/, accessed 14 February, 2024

mining (TDM) for scientific research. This exception allows for the use of copyrighted works without the need for compensation if the use falls within the scope of scientific research and other conditions are met. For example, the DSM Directive specifies that the use of works for text and data mining must be for the sole purpose of scientific research and must be conducted by research organisations, cultural heritage institutions, or other entities with lawful access to the works. Data mining for instance could include activities such as analysing large data sets to discover patterns, trends, or correlations that can be used to improve AI algorithms.[75] This process often involves the use of copyrighted works, such as text, images, or other media, as part of the training data.[76] Additionally, the use must be non-commercial in nature and must not negatively affect the normal exploitation of the works as rightly mentioned by the Berne Convention and TRIPS Agreement. For commercial purposes or if it does not meet the specified conditions for scientific research, it may not fall within the scope of the TDM exception, and compensation or permission from the rights holders may be required.[77] This is to be noted that in the EU, TDM exception is silent on AI data training whether it exempts AI data training from the requirement of compensation or if it is considered a separate category of use.

In the US, the legal framework for text and data mining (TDM) is primarily shaped by copyright law and the fair use doctrine. Unlike the EU, which has specific exceptions for TDM in its Copyright Directive, the US does not have a statutory exception for TDM. Instead, TDM activities are typically analysed under the fair use doctrine, which allows for the limited use of copyrighted works without permission for purposes such as criticism, comment, news reporting, teaching, scholarship, or research. Courts in the US have recognised that TDM for non-expressive purposes, such as extracting facts or data for

[75]Maryna Manteghi, "Overcoming Barriers to Text and Data Mining in the Era of ChatGPT: The Proposed Data Act as a Game-Changer," *GRUR International* 73, no. 1 (2024): 34–44.

[76]Kalpana Tyagi, "Copyright, Text & Data Mining and the Innovation Dimension of Generative AI," *Journal of Intellectual Property Law & Practice*, 2024.

[77] Eleonora Rosati, "An EU Text and Data Mining Exception for the Few: Would It Make Sense?" *Journal of Intellectual Property Law & Practice* 13, no. 6 (2018)

research or analysis, can be transformative and thus may qualify as fair use.[78] However, the outcome of a fair use analysis depends on the specific facts of each case, including the purpose of the use, the nature of the copyrighted work, the amount and substantiality of the portion used, and the effect of the use on the market for the original work akin to section 13(2A) of the Malaysia Copyright Act 1987.

In Malaysia, the legal framework for TDM, specifically in relation to AI data training, is not explicitly addressed in the Copyright Act 1987. However, the Act does provide for fair dealing provisions that could potentially be applied to TDM activities for AI data training influenced by both EU and the US. Fair dealing in section 13(2)(a) of the Copyright Act 1987 allows for the use of copyrighted works for specific purposes including research, private study, criticism, review, and the reporting of current events. A copyright compensation framework for AI training data could be limited to certain uses. For example, the framework could specify that compensation is required for the use of copyrighted works that are used extensively in AI training, or for commercial purposes excluding the use provided under section 13(2)(a) of the Copyright Act 1987. Section 13(2)(a) could be amended to expressly include text and data mining (TDM) activities for AI data training within the scope of fair dealing activities apart from research, private study, criticism, review, or reporting of news or current events. As such, the proposed copyright framework might work if it is designed well to balance the interests of rights holders, users, and AI developers.

[78]Quang, "Training AI," 1407.

CONCLUSION

In conclusion, shaping a possible copyright compensation framework for Malaysia to address AI data training presents both opportunities and challenges. The opportunities lie in creating a framework that encourages the development of AI technologies by providing access to a wide range of copyrighted works. One opportunity is the potential for the framework to stimulate innovation and the development of AI technologies in Malaysia by providing a clear and fair system for accessing copyrighted works. This could attract investment and talent to the country's AI sector, boosting its competitiveness on the global stage.

However, there are challenges to consider. Implementing such a framework requires a careful balance to ensure that rights holders are fairly compensated for the use of their works. Determining the appropriate level of compensation and establishing mechanisms for collecting and distributing royalties can be complex and may require the cooperation of multiple stakeholders. To achieve this balance, the framework should critically provide clear guidelines on the types of works and uses that require compensation, ensuring that rights holders are fairly remunerated without unduly restricting access or hindering innovation. Mechanisms for determining and collecting royalties should be incorporated, potentially through the establishment of a collecting society or licensing body. The framework should also be flexible and adaptable to accommodate future developments in AI technology, ensuring that rights holders continue to receive fair compensation as AI evolves. By addressing these considerations, Malaysia can shape a framework that promotes a fair and sustainable ecosystem for AI development, benefiting both rights holders and the AI industry.

ACKNOWLEDGEMENT

The authors would like to acknowledge the financial support from the Ministry of Higher Education, Malaysia, under the Fundamental Research Grant Scheme FRGS/1/2020/SSO/UKM/01/2.

REFERENCES

Atilla Söğüt, "Dealing with AI-Generated Works: Lessons from the CDPA Section 9(3)." *Journal of Intellectual Property Law & Practice* 19, no. 1 (2024): 43–54.

Andrew W. Torrance and Beth Tomlinson, "Training Is Everything: Artificial Intelligence, Copyright, and Fair Training," *Dickinson Law Review*, forthcoming 2023.

Angeline Yeoh and Christoper Fam, "Moving Forward with AI: How Malaysia Can Grow into an AI-Savvy Nation with the Right Talent and Innovation," *The Star*, August 14, 2023, https://www.thestar.com.my/tech/tech-news/2023/08/14/moving-forward-with-ai-how-malaysia-can-grow-into-an-ai-savvy-nation-with-the-right-talent-and-innovation, accessed June 18, 2024.

Atif Aziz, "Artificial Intelligence Produced Original Work: A New Approach to Copyright Protection and Ownership." *European Journal of Artificial Intelligence & Machine Learning* 2, no. 2 (2023): 3. https://doi.org/10.24018/ejai.2023.2.2.15.

Stanley M, Besen., Sheila. N. Kirby, and Steven C. Salop. "An Economic Analysis of Copyright Collectives." *Virginia Law Review* 78, no. 1 (1992): 383-411.

Blake Brittain, "Getty Images Lawsuit Says Stability AI Misused Photos to Train AI." *Reuters*, February 6, 2023. https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06/, accessed May 29, 2024.

Brian Sanchez, "The Section 115 Mechanical License and the Copyright Modernization Act: The Hardships of Legislating Music Industry Negotiations." *DePaul Journal of Art, Technology & Intellectual Property Law* 17 (2006): 37. https://via.library.depaul.edu/jatip/vol17/iss1/3.

Chawinthorn Watiktinnakorn, Jirawat Seesai, and Chutisan Kerdvibulvech, "Blurring the Lines: How AI Is Redefining

Artistic Ownership and Copyright," *Discover Artificial Intelligence* 3 (2023).

Christina Angelopoulos., and João PedroQuintais. "Fixing Copyright Reform: A Better Solution to Online Infringement." *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law* 10, no. 2 (2019).

Christoph Peukert, "Copyright Levies and Cloud Storage: Ex-Ante Policy Evaluation with a Field Experiment." *Research Policy* 53 (2024): 104918.

Christophe Geiger ., and Vincenzo Iaia. "The Forgotten Creator: Towards a Statutory Remuneration Right for Machine Learning of Generative AI." *Computer Law & Security Review* 52 (2024).

Christopher T. Zirpoli, "Generative Artificial Intelligence and Copyright Law," *Copyright, Fair Use, Scholarly Communication* (2023): 1

Congressional Research Service. "Generative Artificial Intelligence and Copyright Law." *Congressional Research Service*, September 29, 2023. https://crsreports.congress.gov/product/pdf/LSB/LSB10922, accessed February 21, 2024.

David. W. Opderbeck, "Copyright in AI Training Data: A Human-Centered Approach," *Oklahoma Law Review* 76 (2024).

Daniel Rodriguez Maffioli,. "Copyright in Generative AI Training: Balancing Fair Use through Standardization and Transparency." *SSRN*, August 21, 2023. https://ssrn.com/abstract=4579322, accessed June 19, 2024.

Eleonora Rosati, "An EU Text and Data Mining Exception for the Few: Would It Make Sense?" *Journal of Intellectual Property Law & Practice* 13, no. 6 (2018)

Erin Keenan-O'Malley, "'AI' and 'Copyright Infringement' Increasingly Feature in the Headlines, but What Are the Legal Issues Being Debated?" *EIP Newsletter*, UK, December 15, 2023,https://www.aippi.org/news/ai-and-copyright-infringement-increasingly-feature-in-the-headlines-but-what-are-the-legal-issues-being-debated/, accessed 14 February, 2024

Hayden Field, "Anthropic's AI Chatbot Claude Is Posting Lyrics to Popular Songs, Lawsuit Claims." *CNBC*, October 18, 2023. https://www.cnbc.com/2023/10/18/universal-music-sues-anthropic-for-copyright-infringement-over-lyrics.html, accessed June 17, 2024.

FMT Reporters. "Set Up Task Force to Deal with AI Use, Govt Told." *Free Malaysia Today*, February 22, 2024. Retrieved from https://www.freemalaysiatoday.com/category/nation/2024/02/22/set-up-task-force-to-deal-with-ai-use-govt-told/, accessed April 9, 2024.Gerrit De Vynck, "AI Learned from Their Work. Now They Want Compensation." *The Washington Post*, July 16, 2023. https://www.washingtonpost.com/technology/2023/07/16/ai-programs-training-lawsuits-fair-use/, accessed June 2, 2024.

Guadamuz, Andres. "Artificial Intelligence and Copyright." Lecture, University of Sussex, United Kingdom, October 2017.

Haochen Sun, "Redesigning Copyright Protection in the Era of Artificial Intelligence." *Iowa Law Review* 107 (2022): 1213.

Howards B. Abrams, "Copyright's First Compulsory License." *Santa Clara High Technology Law Journal* 26 (2009): 215.

Joost Poort, "Copyright Levies." In *Handbook of the Digital Creative Economy*, edited by R. Towse and C. Handke, 235-246. Edward Elgar, 2013.

Jane C. Ginsburg, . "Fair Use for Free, or Permitted-but-Paid?" *Berkeley Technology Law Journal* 29, no. 3 (2015): 1383-1446.

João Pedro Quintais, "Private Copying and Downloading from Unlawful Sources," *International Review of Intellectual Property and Competition Law* (IIC), Institute for Information Law Research Paper No. 2014-03, Amsterdam Law School Research Paper No. 2014-55, 2015.

Jenny Quang, "Does Training AI Violate Copyright Law?" *Berkeley Technology Law Journal* 36, no. 4 (2021): 1407. https://doi.org/10.15779/Z38XW47X3K.

Jonathan Stempel, "Nvidia Is Sued by Authors Over AI Use of Copyrighted Works," *Reuters*, March 10, 2024, https://www.reuters.com/technology/nvidia-is-sued-by-authors-over-ai-use-copyrighted-works-2024-03-10/, accessed June 217, 2024.

Khaw Lek Tee, "Copyright Law in Malaysia: Does the Balance Hold?" *Journal of Malaysian and Comparative Law* (2004).

Maximiliano Marzetti, "Incorporating Exceptions and Limitations to Copyright Law to Incentivize the Development of Artificial Intelligence in Latin America." *Latin American Law Review* (2022): 19-33.

Marybeth Peters, "The Challenge of Copyright in the Digital Age." *Revista La Propiedad Inmaterial* 9 (2006): 59-67.

Martin Kretschmer, "Private Copying and Fair Compensation: An Empirical Study of Copyright Levies in Europe." *Intellectual Property Office Research Paper* No. 2011/9, University of Glasgow - School of Law, 2011.

Martin Kretschmer, Tzania Margoni, and Pinar Oruç, "Copyright Law and the Lifecycle of Machine Learning Models." *IIC - International Review of Intellectual Property and Competition Law* 55 (2024): 110-138.

Martin Senftleben, "AI Act and Author Remuneration - A Model for Other Regions?" Institute for Information Law (IViR), University of Amsterdam, 2024.

Martin Senftleben, "Generative AI and Author Remuneration." *IIC - International Review of Intellectual Property and Competition Law* 54, no. 4 (2023): 1535-1560.

Michael D. Murray, "Generative AI Art: Copyright Infringement and Fair Use." *SMU Science and Technology Law Review* 26 (2023): 259.

Maryna Manteghi, "Overcoming Barriers to Text and Data Mining in the Era of ChatGPT: The Proposed Data Act as a Game-Changer." *GRUR International* 73, no. 1 (2024): 34–44.

Marlia Zakaria, "MyIPO Bakal Bangkitkan Lagi Isu Penggunaan 'Copyright' AI di Geneva," *Kosmo*, February 15, 2024, https://www.kosmo.com.my/2024/02/15/myipo-bakal-bangkitkan-lagi-isu-penggunaan-copyright-ai-di-geneva/, accessed April 9, 2024.

Nicola Lucchi, "ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems." *European Journal of Risk Regulation* (2023): 1–23.

Noti-Victor Jacob, "Reconceptualizing Compulsory Copyright Licenses." *Stanford Law Review* 72 (2020): 915.

Nurqalby MohdReda, "Malaysia Juga Berdepan Risiko Jenayah AI - Pakar." *Bernama*, February 5, 2024. Retrieved from https://bernama.com/bm/rencana/news.php?id=2268227, accessed April 9, 2024.

Paul Keller, "Protecting Creatives or Impeding Progress? Machine Learning and the EU Copyright Framework," February 17, 2023.

Philip Hacker, "A Legal Framework for AI Training Data—From First Principles to the Artificial Intelligence Act." *Law, Innovation and Technology* 13, no. 2 (2021): 257-301.

Reto M. Hilty, and Valentino Moscon, eds. *Modernisation of the EU Copyright Rules: Position Statement of the Max Planck Institute for Innovation and Competition*. Max Planck Institute for Innovation and Competition Research Paper No. 17-12 (Munich: Max Planck Institute for Innovation and Competition, 2017). https://doi.org/10.17617/2.2470998.

Joan S. Santani, . "Malaysia Coming Up with AI Governance and Code of Ethics." *New Straits Times*, January 9, 2024. https://www.nst.com.my/business/corporate/2024/01/998905/malaysia-coming-ai-governance-and-code-ethics, accessed March 31, 2024.

Stichting De Thuiskopie. *International Survey on Private Copying: Law & Practice 2015*. World Intellectual Property Organization, 2016.

The Star. "AI-produced Content Resembling Works Registered under Intellectual Property a Violation, Says MyIPO." *The Star*, February 15, 2024. https://www.thestar.com.my/news/nation/2024/02/15/ai-produced-content-resembling-works-registered-under-intellectual-properly-a-violation-says

Theodoros Chiou, "Copyright Lessons on Machine Learning: What Impact on Algorithmic Art?" *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law* 10, no. 3 (2019).

Tyagi Kalpana, "Copyright, Text & Data Mining and the Innovation Dimension of Generative AI," *Journal of Intellectual Property Law & Practice*, 2024.

Yolanda King, "The Inadvisability of Nonuniformity in the Licensing of Cover Songs," *Belmont Law Review* 3 (2016): 51

Zhihan Lv, "Generative Artificial Intelligence in the Metaverse Era," *Cognitive Robotics* 3 (2023): 208–217, http://www.keaipublishing.com/en/journals/cognitive-robotics/, accessed March 29, 2024.

Zhuohao Wu, Danwen Ji, Kaiwen Yu and Mohammad Shidujaman, "AI Creativity and the Human-AI Co-creation Model," in *Human-Computer Interaction: Theory, Methods and Tools. HCII 2021. Lecture Notes in Computer Science*, vol. 12762, ed. M. Kurosu (Springer, Cham, 2021).