The Lawfulness of Using Copyrighted Works for Generative AI Training : A Case Study of US Lawsuits against OpenAI and Perplexity AI

Stefany Ismantara[1], Wilma Silalahi[2*]

[1,2] Universitas Tarumanagara, Indonesia

correspondence email : wilmasilalahi@fh.untar.ac.id

Abstract: Copyright protection in Indonesia is governed by Law No. 28 of 2014 (Law 28/2014), encompassing moral and economic rights. This law imposes limitations on the use of works for education, law enforcement, or technological development as long as such use does not harm the legitimate interests of the rightholder. The development of generative AI (GAI) poses challenges in determining the legality of using copyrighted works for GAI training. This study examines copyright regulations concerning GAI through normative, conceptual, and comparative legal approaches, including case studies on lawsuits against OpenAI and Perplexity AI. The findings indicate that the legality of using copyrighted works depends on the data input process and output (responses). Data scraping is considered an economic right of the rightholder, classified as reproduction under Article 9 of Law 28/2014. If such acts are conducted without the rightholder's consent and for commercial purposes, they are deemed unlawful under Indonesia's current copyright law. GAI outputs may also infringe copyright if: (1) the source is not cited, violating Article 7 on copyright management information; (2) substantial portions of the work are reproduced, violating the rightholder's economic rights under Article 9; or (3) the work is distorted in a way that harms the rightholder's honor, infringing on moral rights under Article 5. To accommodate AI development, specific regulations integrating AI transparency principles outlined in SE Kominfo 9/2023 are required. These regulations could include obligations for AI companies to release summaries of training datasets, including the EU AI Act that define the responsibilities of AI developers and users, and provide disclaimers regarding AI's limitations. Regarding the fulfillment of rightholders' economic rights, a non-exclusive blanket license through Collective Management Organizations (CMOs) as stipulated in Permenkumham 15/2024 is necessary. These regulations should be synchronized with related policies to establish legal certainty that adapts to technological advancements.

Keywords: Copyright; Copyrighted Works; Generative Artificial Intelligence

Abstrak: *Perlindungan hak cipta di Indonesia diatur dalam UU Nomor 28 Tahun 2014 (UU 28/2014), mencakup hak moral dan hak ekonomi. UU ini memberikan pembatasan penggunaan ciptaan untuk pendidikan, penegakan hukum, atau pengembangan teknologi selama tidak merugikan kepentingan wajar pencipta. Perkembangan generative AI (GAI) menimbulkan tantangan dalam menentukan keabsahan penggunaan ciptaan untuk pelatihan GAI. Penelitian ini mengkaji regulasi hak cipta terkait GAI melalui pendekatan yuridis normatif,*

konseptual, dan komparatif, dengan studi kasus gugatan terhadap OpenAI dan Perplexity AI. Hasil penelitian ini menunjukkan bahwa keabsahan penggunaan ciptaan tergantung pada proses input data dan output (jawaban). Data scraping dianggap sebagai hak ekonomi Pencipta untuk menggandakan Ciptaannya sesuai dengan Pasal 9 UU 28/2014. Apabila perbuatan tersebut dilakukan tanpa izin Pencipta dan dilakukan secara komersial, maka tindakan tersebut menjadi tidak sah dalam hukum hak cipta di Indonesia saat ini. Output AI juga berpotensi melanggar hak cipta jika: (1) tidak mencantumkan sumber, melanggar Pasal 7 tentang informasi manajemen hak cipta; (2) menyalin bagian substansial ciptaan, melanggar hak ekonomi pencipta di Pasal 9; atau (3) mendistorsi ciptaan hingga merugikan kehormatan pencipta, melanggar hak moral di Pasal 5. Dalam mengakomodasi perkembangan AI, diperlukan regulasi khusus yang mengintegrasikan prinsip transparansi AI dalam SE Menkominfo 9/2023. Substansi regulasi ini dapat mencakup kewajiban perusahaan AI untuk merilis ringkasan dataset pelatihan, mencantumkan substansi dari UU AI Uni Eropa yang mengatur batasan-batasan tanggung jawab antara perusahaan pengembang AI dengan pengguna, serta memberikan disclaimer terkait keterbatasan AI. Berkaitan dengan pemenuhan hak ekonomi Pencipta, perlu diadakan blanket license yang bersifat non-eksklusif melalui Lembaga Manajemen Kolektif (LMK) sebagaimana diatur Permenkumham 15/2024. Regulasi-regulasi tersebut perlu disinkronisasikan dengan regulasi terkait lainnya agar menciptakan kepastian hukum yang adaptif terhadap perkembangan teknologi.

Kata Kunci : *Hak Cipta; Ciptaan; Generative Artificial Intelligence*

INTRODUCTION

The concept of copyright was introduced in the 15th century in response to publishers' need for protection against duplication of copyrighted works. The urgency behind the development of copyright law lies in the Commercialization aspect of copyrighted works.[1] The first Copyright Law, the Statute of Anne 1710, established the exclusive right to copy and distribute copyrighted works.[2] In Indonesia, the national copyright law was first enacted after Indonesia's independence in 1982 and then underwent several amendments until it became the current Copyright Law, namely Law Number 28 Year 2014 on Copyright ("Law 28/2014").[3] Law 28/2014 defines Copyrighted Works as a creative work in the fields of science, art, and literature that is produced by inspiration, ability, thought, imagination, dexterity, skill, or expertise expressed in a tangible form.[4] While Copyright is an exclusive right owned by the copyright owner that arises automatically based on the declarative principle after a work is realized in real form.[5]

---

[1] Nainggolan, Bernard. *Perlindungan Hukum Kekayaan Intelektual di Era Digital.* Yogyakarta : Publika Global Media, 2021. p.12-13.

[2] Triatmojo, Firmandanu et al. *Perlindungan Hak Cipta Lagu Komersil.* Jawa Tengah : NEM, 2021. p.9

[3] Rizkia, Nanda Dwi & Fardiansyah, Hardi. *Hak Kekayaan Intelektual Suatu Pengantar.* Bandung : Penerbit Widina, 2022. p.31.

[4] Indonesia, Undang-undang Nomor 28 Tahun 2014 tentang Hak Cipta (Lembaran Negara Republik Indonesia Tahun 2014 Nomor 266, Tambahan Lembaran Negara Republik Indonesia Nomor 5599), Pasal 1 ayat (3). https://peraturan.bpk.go.id/Details/38690.

[5] Pasal 1 ayat (1), *Loc.Cit.*

In Law 28/2014 there are restrictions on copyright. In terms of copyright restrictions, a person may use copyrighted material for the purposes of education, information dissemination, law enforcement or technology development, provided that they do not directly infringe copyright by taking a substantial part of the work (known as fair use).[6] These restrictions are intended to encourage the advancement of science and technology. In relation to technological progress, along with the times, technology has developed to the point that it can imitate human intelligence, which is called artificial intelligence (AI). AI is a computer system-based technology that allows the system to perform activities that usually require human intelligence.[7] One of the ways AI learns algorithms is through machine learning (ML). Machine learning is a branch of computer science and AI that studies how to create systems and algorithms that can learn and evolve independently.[8]

The early 2000s was a pivotal moment in the rapid development of machine learning. Various ML techniques that can analyze large amounts of data make it easy to find patterns in data, gain insights, and automate certain tasks. Then, around 2010, ML capabilities grew further with the arrival of deep learning (DL), which contributed to the development of online search engines, voice recognition, and self-drive cars. DL capabilities further developed until Generative AI (GAI) emerged.[9] GAI is an AI that can create new content such as text, images, video, and even audio that has never existed before.

Advances in GAI technology have brought significant benefits in various sectors, from education to the creative industry. However, behind this potential, problems arise related to copyright protection. The reason is that GAI is trained using training datasets. Through these datasets, GAI can understand patterns that govern complex relationships and relevance between various data and sample from that understanding, which is then used as a reference to create new content.[10] The problem is that the datasets used to train GAI generally come from big data or public data sets that come from the thoughts or intellectual work of various parties.[11]

As written earlier, Law 28/2014 has provided protection for the use of works that are still within the copyright restrictions stipulated in the law, including one for technological development. However, this raises questions when the use of a work is done without permission to develop AI and commercial benefits are derived from the development. This touches on two aspects, namely related to the input process of the work as GAI training data,

---

[6] Kanti Rahayu, "Kajian Hukum Tentang Penerapan Fungsi Sosial Pada Hak Cipta," Diktum: Jurnal Ilmu Hukum 9, no. 2 (22 Agustus 2022): 154, https://doi.org/10.24905/diktum.v9i2.92.

[7] Zhang, Caiming & Lu, Yang. Study on Artificial Intelligence: The State of The Art and Future Prospects. Journal of Industrial Information Integration, Vol. 23, Artikel 100224. https://doi.org/10.1016/j.jii.2021.100224.

[8] Amazon. What is Natural Language Processing (NLP)?. Diakses dari https://aws.amazon.com/what-is/nlp/.

[9] Jamaaluddin, MM & Sulistyowati, Indah. Buku Ajar Kecerdasan Buatan (Artificial Intelligence). Sidoarjo : Umsida Press, 2021. p.29.

[10] François Fleuret. The Little Book of Deep Learning. Universite de Geneve, 2024. p. 12.

[11] Kan Hyden, "AI, Norms, Big Data, and the Law," Asian Journal of Law and Society 7, no. 3 (1 November 2020): 409–436, https://doi.org/10.1017/als.2020.36.

and the output process of GAI containing information related to the work.

In this landscape, various lawsuit cases against GAI companies in the United States have provided concrete illustrations of the problem. These include lawsuits against OpenAI, Stability AI, Meta Platforms, and others.[12] However, this scientific article will focus on the GAI text model, mainly by examining the validity of the use of training data in ChatGPT and Perplexity AI.

Previous studies have explored the theoretical and policy aspects of the polemic between AI development and copyright protection. For example, the concept of fair use in some jurisdictions has long been considered as one of the solutions to balance legal protection with the needs of technological innovation. However, in practice, there is still a gap between the theoretically formulated legal principles and their application in real legal disputes, where the law is substantially less able to keep pace with the needs of technological development and the fulfillment of people's rights as creative individuals.

However, there is no study that focuses on the validity of the use of training data by GAI text models, especially in the context of the input and output produced by ChatGPT and Perplexity AI as will be described in the discussion in this research article. This study seeks to bridge the existing gap by examining concrete cases that can provide answers to the validity of the use of copyrighted works for GAI training by analyzing a comparison of copyright and AI law in the United States and Indonesia through a case study of the Raw Story, Inc. & AlterNet, Inc. lawsuit against OpenAI, Inc. and the Dow Jones & Company, Inc.; NYP Holdings, Inc. lawsuit against Perplexity AI, Inc. This scientific study aims to explore alternative regulations that can be formed as a middle ground for copyright protection and the interests of technological development. This research is expected to be a theoretical reference as well as a practical recommendation that can be used by policy makers, researchers, and legal practitioners in contributing new insights that enrich legal discourse related to copyright and AI technology.

METHOD

This research uses a normative juridical approach to analyze theories, concepts, and legislation relevant to the research topic. The approaches used include a statutory approach by analyzing applicable regulations related to the legal issues at hand, a conceptual approach based on the development of doctrines and views in legal science, and a comparative approach that compares regulations in Indonesia and internationally to find relevant legal reform solutions to the legal vacuum related to GAI and copyright.[13]

---

[12] Zeynep Ülkü Kahveci, "Attribution problem of generative AI: a view from US copyright law," *Journal of Intellectual Property Law and Practice* 18, no. 11 (1 November 2023): 796–807, https://doi.org/10.1093/jiplp/jpad076.

[13] Djulaeka & Rahayu, Devi. Buku Ajar: Metode Penelitian Hukum. Surabaya : Scopindo Media Pustaka, 2020. pp. 88. https://books.google.co.id/books?id=aIrUDwAAQBAJ&printsec=copyright&redir_esc=y#v=onepage&q&f=false.

The data in this research is collected through literature study that relies on secondary data, including primary legal materials such as laws and regulations, secondary legal materials such as relevant literature, jurists' views, journals, and papers, tertiary legal materials such as legal dictionaries and encyclopedias, as well as non-legal materials such as Indonesian dictionaries; books, papers, and journals regarding the study of AI, along with interviews with experts both academics and practitioners to help gain a more comprehensive understanding of the legal issues discussed.[14]

The analysis was conducted qualitatively by collecting, selecting, and interpreting legal materials from various literature sources and related events. This process uses the deductive method, starting from general premises or major propositions, then formulating more specific minor premises, resulting in specific conclusions or prescriptions. The conclusions are organized descriptively to provide a comprehensive understanding of the legal issues discussed.[15]

DISCUSSION

A. Lawfulness of Using Copyrighted Works for Generative AI Training (ChatGPT & Perplexity AI Case Study)

OpenAI is a company that develops ChatGPT application which is a GAI Chatbot technology. As the name implies, ChatGPT is developed using Generative Pre-trained Transformer technology that can transform the data used to train it into transformative content.[16] Since its inception in 2022, the OpenAI company has received numerous copyright infringement suits by the original owners of the content it uses to program ChatGPT. One of the most recent suits filed against the company was filed by Raw Story Media, Inc and AlterNet Media, Inc as of February 28, 2024.[17] Both are journalistic publishing companies operating in the United States.

Raw Story and AlterNet base their claims on violations of Title 17 of the United States Code ("US Copyright Act") and the Digital Millennium Copyright Act (DMCA). They base the violations on: 1) Section 1202(b)(1) of the US Copyright Act which prohibits the intentional removal or alteration of copyright management information (CMI) without the copyright owner's permission or without being legally authorized; and 2) Section 106(1), (2) of the US Copyright Act which provides for the exclusive right of the copyright owner to make and grant permission to reproduce the copyrighted work in copies or sound recordings and to prepare derivative works based on the copyrighted work.

---

[14] "Juliardi, Budi et al. Metode Penelitian Hukum. Sumatera Barat : CV. Gita Lentera, 2023. p. 43. https://books.google.co.id/books?id=vyXbEAAAQBAJ&printsec=frontcover&hl=id#v=onepage&q&f=false. ," t.t.
[15] *Ibid.*, p. 84.
[16] Viriya Taecharungroj, "'What Can ChatGPT Do?' Analyzing Early Reactions to the Innovative AI Chatbot on Twitter," *Big Data and Cognitive Computing* 7, no. 1 (1 Maret 2023), https://doi.org/10.3390/bdcc7010035.
[17] Emilia David. The Intercept, Raw Story, and AlterNet sue OpenAI and Microsoft. Feb 29, 2024. https://www.theverge.com/2024/2/28/24085973/intercept-raw-story-alternet-openai-lawsuit-copyright.

ChatGPT has no independent knowledge of the information provided in its responses. Instead, in providing services to customers who have paid to use its services, ChatGPT repackages papers created by various parties as its output. Thus, when providing responses, ChatGPT gives the impression that it is an all-knowing and "intelligent" source of information, which is inconsistent with the reality that the responses are often based on copyrighted journalistic works that ChatGPT has merely copied.[18]

In copyright law in Indonesia, these regulations are listed in the moral rights owned by the copyright owner as stipulated in Article 5 Paragraph (1) of Law 28/2014. The article explains that moral rights are rights that are eternally attached to the Copyright Owner, one of which includes putting or not putting his name on copies that are used publicly. Article 6 of Law 28/2014 also regulates Copyright management information and Copyright electronic information, which is followed by the provision in Article 7 that the information may not be removed, altered, or destroyed. Copyright restrictions stipulated in Article 43 of Law 28/2014 also state that the taking of actual news, either in whole or in part from news agencies, broadcasting institutions, and newspapers or other similar sources provided that the source must be stated in full. Thus, in copyright law in Indonesia the act of abolishing CMI can be said to be unauthorized or illegal.

Furthermore, the plaintiff claimed that OpenAI had also taken unauthorized copies of works published by Raw Story & AlterNet on the internet and used them to train ChatGPT, after which the output generated by ChatGPT contained information from Raw Story & AlterNet's work. The claim is corroborated by data from the award-winning website Copyleaks, which states that nearly 60% of the responses provided by OpenAI's GPT-3.5 product contain some form of plagiarized content, and over 45% contain text identical to pre-existing content. While the plaintiffs were unable to prove the similarity or the specific substantial portions that were 'plagiarized' by ChatGPT, this is because OpenAI has kept secret the specific content used to train all versions of ChatGPT.

In Law 28/2014, the act of reproducing copies of Copyrighted Works is also prohibited. It is clear that the exclusive rights of the Copyright Owner are in the form of moral rights and economic rights over Copyrighted Works. The economic rights are regulated in Article 9, including the right to duplicate Copyrighted Works in all its forms, adaptation, arrangement, transformation of Copyrighted Works, and distribution of Copyrighted Works or their copies. Any person who exercises such economic rights must obtain the permission of the Copyright Owner or Copyright Holder. Such permission is generally implemented by entering into a license agreement. It is stipulated that any Person without the permission of the Copyright Owner or Copyright Holder is prohibited from reproducing and/or commercially using the Copyrighted Work.[19]

In relation to licenses, the Regulation of the Minister of Law and Human Rights

---

[18] Amerika Serikat, Surat Gugatan OpenAI v. Raw Story Media, Inc dan AlterNet Media, Inc.

[19] Punik Triesti dkk., "Web Scraping dalam Aplikasi ChatGPT oleh Chatbot Berbasis Artificial Intelligence Berdasarkan Undang-Undang Nomor 28 Tahun 2014 Tentang Hak Cipta," Jurnal Demokrasi dan Ketahanan Nasional |, vol. 3, no. 2 (2024): 118, DOI: https://doi.org/10.20961/souvereignty.v3i2.1620.

Number 15 of 2024 concerning Management of Royalties on Secondary Use Licenses for Copyright of Books and/or Other Written Works (hereinafter referred to as "Permenkumham 15/2024") has provided legal certainty for Copyright Owners of books or other written works in obtaining royalties for the reproduction of books and/or other written works, whether the reproduction is done digitally or non-digitally. Article 5 of MOLHR 15/2024 states that Secondary Users of copyrighted books and/or other written works include private businesses that carry out document duplication activities, electronic system providers, and artificial intelligence developers.[20] Therefore it becomes clear that AI developers including GAI Chabot, according to the regulation, are classified as Secondary Users of Copyrighted Works. Article 17 paragraph (1) of Permenkumham 15/2024 states that the secondary use of Copyrighted Works on Books and/or Other Works of Writing can be done without asking for prior permission from the Copyright Owner and/or Copyright Holder by paying a fee through the Collective Management Institution (LMK) in the field of Books and/or Other Works of Writing. Article 18 paragraph (1) of Permenkumham 15/2024 states that one form of secondary use of the Copyrighted Work is the duplication and/or distribution of the Copyrighted Work of Books and/or Other Works by retrieving data from a large number of works on the internet (web scraping). Based on these articles, it is clear that in Indonesian copyright law, the duplication and distribution of Copyrighted Works as a GAI training dataset is not legal if it is done for commercial purposes and without entering into a license agreement, either with the Copyright Owner directly or through the LMK where the Copyright Owner is registered as a member.

Perplexity, on the other hand, is a generative AI company that claims to provide accurate and up-to-date news and information to its users on a single platform.[21] The information, like generative AI in general, is taken from data spread across the internet, including journalistic content. However, unlike generative AI like ChatGPT, Perplexity AI includes citations for the answers or responses it provides.[22]

However, Perplexity AI has not escaped a lawsuit from the owners of the content it copied as its training data. On October 21, 2024, Dow Jones, NYP Holdings, and its parent company, News Corp, which is one of the publishers that has produced many well-known works, including The Wall Street Journal and the New York Post which are two of the most widely circulated newspapers in the United States (hereinafter 'Plaintiff'), filed a lawsuit

[20] HukumOnline. Kemenangan bagi Pemegang Karya Tulis dengan Hak Cipta! Batasan Penggunaan Lisensi Sekunder Ditetapkan 10%. (2024, Juli 5). Diakses dari https://pro.hukumonline.com/a/lt6687682cc3d63/kemenangan-bagi-pemegang-karya-tulis-dengan-hak-cipta-batasan-penggunaan-lisensi-sekunder-ditetapkan-10.

[21] Tira Nur Fitria, "Using ChatBot-Based Artificial Intelligence (AI) for Writing an English Essay: The Ability of ChatGPT, Perplexity AI, and ChatSonic," Journal of Language Intelligence and Culture 6, no. 2 (26 Agustus 2024): 103–128, https://doi.org/10.35719/jlic.v6i2.139.

[22] Shtykalo, O & Yamnenko, I. ChatGPT and Other AI Tools for Academic Research and Education. Digital Ecosystems: Interconnecting Advanced Networks with AI Applications, Vol. 1198, (2024). DOI: https://doi.org/10.1007/978-3-031-61221-3_29.

against Perplexity AI, Inc (hereinafter 'Perplexity' or 'Defendant').

The plaintiffs base the violation on the following legal grounds: 1) Article 106 of the US Copyright Act - Perplexity's copying of Plaintiff's Copyrighted Works as "Input" for its Retrieval-augmented Generation (RAG) Index. In this case, Perplexity has committed infringement because it has unlicensed and illegally copied a large number of Plaintiff's works as inputs into its RAG index; and 2) Section 106 of the US Copyright Act - Perplexity's Copying of Plaintiff's Copyrighted Works to Produce "Output" to answer User Queries; and 3) Section 1125 of Title 15 of the United States Code (Commerce and Trade) - Improper Reference and Trademark Infringement.

According to the Plaintiff, Perplexity has committed a large number of illegal copying of the plaintiff's copyrighted works and diverted customers and deprived the plaintiff of revenue. Perplexity was programmed with an 'answer engine'. The answer engine massively copied copyrighted news, analysis and opinion content as input into its database. Perplexity then used the copyrighted content to generate responses to user questions that were intended to serve as a substitute for news and other information websites. The plaintiffs argue that Perplexity claims that the answers to user queries are so reliable that users can "Skip the Link" to the original publisher and rely entirely on Perplexity for their news and analysis needs. Perplexity AI has thus deprived the Plaintiff of a crucial source of revenue.

Aside from the input process, the output of Perplexity's products also infringed Plaintiff's copyrights in an unauthorized manner. Perplexity's "answers" to user queries often included full or partial verbatim copying of Plaintiff's news, analysis, and opinion articles. Users could access verbatim copies of Plaintiffs' content more frequently by purchasing a subscription to Perplexity's premium service, "Perplexity Pro." Perplexity also frequently converted Plaintiff's copyrighted articles into paraphrases or summaries of such copyrighted works that also served as substitutes for accessing Plaintiff's copyrighted works on Plaintiff's own and/or licensed websites. The Plaintiff considers that the use of the Plaintiff's copyrighted content to produce such a substitute is not a fair use, as it has harmed the reasonable interests of the Copyright Owner. The claim is reinforced by the Plaintiff's experience of detecting almost no traffic to their website from the links included on Perplexity's answers, despite Perplexity having received approximately 250 million queries per month.[23]

The lawsuit is similar to the Article of the lawsuit against OpenAI described earlier, namely Article 106 of the US Copyright Act. As mentioned, the act of copying is not legal under Law 28/2014 because commercial copying and distribution of Copyrighted Works is the exclusive right of the Copyright Owner and therefore requires a license agreement. However, there is a slight difference between the two cases, which is due to the different ways in which ChatGPT and Perplexity AI produce output. ChatGPT is transformational and does not show references in its output, while Perplexity AI shows references accompanied

---

[23] Amerika Serikat, Surat Gugatan Dow Jones & Company, Inc.; NYP Holdings, Inc., vs Perplexity AI, Inc.

by links that direct users to the original source. The inclusion of such references, in Indonesian copyright law, can be an umbrella that protects the use of Copyrighted Works. This is regulated in Article 48 of Law 28/2014 which states that the duplication, broadcasting, or communication of Copyrighted Works for informational purposes that mention the source and the name of the Copyright Owner in full is not considered an infringement of Copyright provided that the Copyrighted Work is in the form of articles that have been published in both print and electronic media except for copies provided by the Copyright Owner, or related to broadcasting or communication of a Copyrighted Work; reports of actual events or brief excerpts from the Copyrighted Work seen or heard in certain situations; and scientific works, speeches, lectures, or similar Copyrighted Works delivered to the public. Then Article 43 of Law 28/2014 also states that the retrieval of actual news, either in whole or in part provided that the source must be fully mentioned as well as the Copyrighted Work and dissemination of Copyright content through information and communication technology media that is non-commercial and/or benefits the Copyright Owner or related parties, or the Copyright Owner expresses no objection to the Copyrighted Work and its dissemination, is not considered copyright infringement. However, here there is no clarity regarding the validity and consequences if a Copyrighted Work is duplicated and communicated by including references, but at the same time the behavior results in commercial benefits for the perpetrator.

When viewed from Article 18 paragraph (3) of Permenkumham 15/2024, secondary use of Book and/or Other Works of Writing is a copy and/or distribution that does not exceed 10% (ten percent) of the Book and/or Other Works of Writing by secondary users as referred to in Article 7; or b. does not harm the reasonable interests of the Copyright Owner and/or Copyright Holder. Then in paragraph (4), it is explained that the secondary use of Copyright of written works contained in magazines, newspapers, or periodicals may not cover the whole of 1 (one) article.[24] The regulation provides clarity regarding the portion that can be taken for duplication and/or distribution of Copyrighted Works. However, there is an overlap with Law 28/2014 in terms of actual news retrieval, where in Article 43 of Law 28/2014 such retrieval is allowed either in whole or in part as long as it includes the complete source, while in Article 18 paragraph (4) of Permenkumham 15/2024 it is stated that such retrieval cannot be taken in its entirety. This overlap has resulted in legal uncertainty regarding the use of Copyrighted Works in the form of actual news. However, with the existence of these regulations, it is clear that the duplication and/or distribution of Copyrighted Works with the provision of references can still be sued if it harms the reasonable interests of the Copyright Owner. Therefore, the inclusion of references in this case cannot release the full responsibility of the user of the Copyrighted Work, because the inclusion of references only fulfills the exclusive rights in the form of moral rights. Regarding economic rights, it can still be debated in court if the Copyright

---

[24] HukumOnline., *Loc.Cit.*

Owner feels that his economic rights have been harmed by the secondary use.[25]

In addition to using Plaintiff's copyrighted works to develop substitute products that copied or imitated Plaintiff's original content, Perplexity also harmed Plaintiff's trademark by falsely attributing certain content that was never authored or published by Plaintiff. Sometimes, when Perplexity users inquire regarding the content of the Plaintiff's content, Perplexity responds with incorrect information. As such, the Plaintiff's well-known trademark, which has a reputation for credible accuracy, was damaged and assimilated with hallucinations created by Perplexity's "answer machine", thereby spreading disinformation to the public to believe that the hallucinations were genuine news published by the Plaintiff.[26] In Law 28/2014, such actions can be classified as a violation of the moral rights of the Copyright Owner as stipulated in Article 5, namely the right to defend their rights in the event of distortion of Copyrighted Works, mutilation of Copyrighted Works, modification of Copyrighted Works or things that are detrimental to their personal honor or reputation.

There is a difference between the way ChatGPT and Perplexity AI work, which results in a difference to the basis of the lawsuit as described. OpenAI was sued for abolishing CMI by not including citations, even though the work produced was transformative, i.e. new content. This means that the answers produced by OpenAI are not solely taken from one or two sources, but from thousands of datasets (big data) that are automatically retrieved (text and data mining) by algorithms with web crawling.[27] Afterwards, the data is assimilated and produces new, transformative content. In this case, the plaintiff could not prove the similarity of ChatGPT's output to its own work and could only prove it through a general plagiarism test. Meanwhile, Perplexity AI is programmed to include citations to the output it provides at the user's request, therefore the output provided tends to point to a particular source, thus the percentage of similarity is higher and cannot be said to be transformative. Perplexity states that its AI does not act as a web crawler, but rather as a tool to help users retrieve and process the information they request.[28]

The issue lies in the fact that their practice of providing answers with citations is being challenged on the grounds that it portrays the AI as "reliable," potentially encouraging users to rely on Perplexity's output without clicking on the cited links. Additionally, there is a claim of trademark infringement, which highlights that the AI's

[25] Faidatul Hikmah, et al. "Perlindungan Hak Ekonomi Bagi Pemilik Hak Cipta Dalam Perspektif Hukum Kekayaan Intelektual Di Indonesia," Jurnal Pendidikan dan Dan Konseling vol.5 no.2 (2023): 2254-2260. DOI:https://doi.org/10.31004/jpdk.v5i2.13503.

[26] Deike, Michael. 2024. "Evaluating the Performance of ChatGPT and Perplexity AI in Business Reference." Journal of Business & Finance Librarianship 29 (2): 125–54. doi:10.1080/08963568.2024.2317534.

[27] Kalpana Tyagi, "Copyright, text & data mining and the innovation dimension of generative AI," Journal of Intellectual Property Law and Practice vol, 19 no. 7 (1 Juli 2024): 557–570, https://doi.org/10.1093/jiplp/jpae028.

[28] Aamo Iorliam dan Joseph Abunimye Ingio, "A Comparative Analysis of Generative Artificial Intelligence Tools for Natural Language Processing," Journal of Computing Theories and Applications vol. 1, no. 3 (26 Februari 2024): 313, https://doi.org/10.62411/jcta.9447.

operational mechanisms can result in the delivery of inaccurate responses, making the occurrence of misinformation in its output not uncommon.[29] The disinformation is neither intentional on the part of the programmer nor the AI company. This is because, under the principles of Machine Learning (ML), AI systems are designed to learn, evolve, and operate autonomously. Consequently, the responses generated by AI are independent of direct human control.[30]

The similarity between the two lawsuits is that they are both based on the violation of Article 106 of the US Copyright Act related to unauthorized copying of works. In essence, the copying and/or distribution of a Copyrighted Work that is not done commercially and does not harm the reasonable interests of the Copyright Owner is valid under copyright law, both the US Copyright Act and Law 28/2014. However, when such use is done commercially and without the authorization of the Copyright Owner, the act becomes unauthorized and can be classified as copyright infringement.[31] The copyright restrictions set out in Article 43 to Article 51 of Law 28/2014 only allow the use of material that does not directly infringe copyright by taking a substantial part of a work (known as fair use). Law 28/2014 does not provide clear factors related to fair use.

However, Section 107 of the US Copyright Act mentions factors to be considered in determining whether the use of a work in a particular case is a fair use, including the purpose and character of the use, whether the use is commercial or for non-profit educational purposes; the nature of the copyrighted work; the amount and substantiality of the part used in relation to the copyrighted work as a whole; and the impact of the use on the potential market or value of the copyrighted work.[32]

Analyzing the lawsuits filed against OpenAI, Inc. and Perplexity AI, Inc., it is evident that both entities are private companies deriving commercial benefits from the development of their Artificial Intelligence (AI) programs. During the development process, these companies rely on inputs in the form of training datasets, which are subsequently reflected in the outputs produced by the AI systems. These outputs create a market that indirectly generates economic advantages for the companies, both through the growing number of users and the significant revenue derived from the purchase of "premium" services by their users. Thus, when looking from the perspective of copyright law, the act has touched the 'reasonable interests' of the Copyright Owner and thus becomes invalid.

---

[29] Chiara Longoni et al., "News from Generative Artificial Intelligence Is Believed Less," dalam ACM International Conference Proceeding Series (Association for Computing Machinery, 2022), 104, https://doi.org/10.1145/3531146.3533077.

[30] Manali Shukla et al., "A Comparative Study of ChatGPT, Gemini, and Perplexity," International Journal of Innovative Research in Computer Science and Technology vol. 12, no. 4 (Juli 2024): 13, https://doi.org/10.55524/ijircst.2024.12.4.2.

[31] Ninda Alfani et al., "Implementasi UU Nomor 28 Tahun 2014 dalam Perlindungan Hak Cipta di Era Digital:, Journal of Administrative and Social Science, vol. 4 no. 1 (2022): 23-36. DOI: https://doi.org/10.55606/jass.v4i1.

[32] Raden Radisa Difa Devina dan Tatty Aryani Ramli, "Penggandaan dan Pengumuman Karya Cipta E-Book berdasarkan Prinsip Fair Use Ditinjau dari Hukum Positif," Bandung Conference Series: Law Studies vol. 2 no. 1, (2022): 462. https://doi.org/10.29313/bcsls.v2i1.890.

The enactment of LMK Karya Tulis through Permenkumham 15/2024 indirectly confirms that the fulfillment of moral rights alone against the commercial use of Copyrighted Works in this case is not sufficient, even though the use is only in the form of web scraping. So the invalidity of the use of Copyrighted Works for GAI training in the context of interpretation of copyright law in Indonesia becomes clear.

However, an interdisciplinary approach to the validity of using Copyrighted Works as GAI training data is required. In this case, it is necessary to understand how the GAI works and the legal subjects who play a role in a series of events and/or actions that result in the violation of the law. Fundamentally, Artificial Intelligence (AI) is a program developed by programmers with clearly defined instructions, where the data utilized is sourced from cyberspace via the internet (big data) based on the commands provided. In this context, three key elements play a critical role: programmers, big data, and users. Programmers are responsible for designing AI systems to navigate and process big data, while big data refers to a collection of information that has been made publicly accessible on the internet (published: made available to the public).[33] The party that controls the use of such data is the user. Therefore, it cannot be asserted that either the AI or the programmer possesses the intent to commit or avoid copyright infringement, as the user maintains control over the AI's application. However, the AI development company holds the responsibility to ensure that the training data is acquired and utilized in compliance with applicable laws. While the programmers merely develop tools that are not directly tied to the final output, the responsibility for the use of AI-generated results or outputs lies entirely with the end users.[34]

Thus, when looking at the elements of infringement in Law 28/2014, the acquisition and use of Copyrighted Works as input data for GAI training is illegal because it has harmed the reasonable interests of the Copyright Owner. However, with regard to the output, due to the nature of AI acting as a tool, it is not appropriate to impose the responsibility entirely on the AI programmer or AI company.[35] However, AI companies still have a social responsibility as companies that facilitate the use of the data obtained and the content and/or answers produced.[36] This needs to be further regulated in the legal framework regarding artificial intelligence.

Although there is no specific regulation on AI in Indonesia that is included in the hierarchy of legislation, its regulation can follow existing regulations regarding electronic systems, because AI can be categorized as an electronic system. Regulations that can

---

[33] Adam J. Andreotta, Nin Kirkham, dan Marco Rizzi, "AI, big data, and the future of consent," AI and Society vol. 37, no. 4 (1 Desember 2022): 1715–1728, https://doi.org/10.1007/s00146-021-01262-5.

[34] Mark Coeckelbergh, "Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability," Science and Engineering Ethics vol 26, no. 4 (1 Agustus 2020): 2060, https://doi.org/10.1007/s11948-019-00146-8.

[35] Lucchi, Nicola. "ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems." European Journal of Risk Regulation, (2023):1–23. DOI: 10.1017/err.2023.59.

[36] Mark Anthony Camilleri, "Artificial intelligence governance: Ethical considerations and implications for social responsibility," Expert Systems vol. 41, no. 7 (1 Juli 2024), https://doi.org/10.1111/exsy.13406.

accommodate it include Government Regulation No. 71/2019 on the Implementation of Electronic Systems and Transactions, and Minister of Communication and Information Technology Regulation No. 5/2020 on Private Scope Electronic System Operators and its amendment through Minister of Communication and Information Technology Regulation No. 10/2021.[37]

B. Alternative Regulations to the Development of AI Technology and Copyright

The G20 Artificial Intelligence Principles agreed at the G20 Summit meeting in Osaka in 2019, set out principles such as inclusive growth, transparency, security, and accountability. Indonesia, towards the G20 Summit, took into account the development of AI based on these principles and integrated it with the values of Pancasila through the ethics of artificial intelligence created through the National Strategy for Indonesian Artificial Intelligence 2020-2045 by the Agency for the Assessment and Application of Technology (BPPT). Then the substance of the BPPT was translated into the Circular Letter of the Minister of Communication and Information Technology Number 9 of 2023 (SE Menkominfo 9/2023). This Circular Letter stipulates nine ethical values that should be considered in the development of AI technology such as inclusivity; humanity; security; democracy; transparency; credibility and accountability; protection of personal data; environmental development and sustainability; and intellectual property rights.[38]

In the United States, the "Risk Management Profile for Artificial Intelligence and Human Rights" was also set up in July 2024 as a practical guide to the development and use of AI in a consistent manner that respects human rights, including with respect to safety, security and resilience, transparency and accountability, privacy, and fairness and bias. The principle that stands out in relation to the use of Copyrighted Works as GAI training data is the principle of transparency.[39]

Regulation in the form of an Act regarding GAI in the United States first appeared in the State of Utah, on May 1, 2024 through the "AI Policy Act". The Act emphasizes transparency obligations for companies developing generative AI in Utah. The Act requires GAIs to inform users who interact with the GAI about its status as a GAI. The disclosure must be clearly and explicitly stated. In another US state, California, on September 19, 2024, Bill SB-942 was adopted: California AI Transparency Act. The Act also emphasizes the obligation of AI companies, especially GAIs, to uphold the principle of transparency in their development.[40]

---

[37] Putra, G., Taniady, V., & Halmadiningrat, I. Tantangan Hukum: Keakuratan Informasi Layanan AI Chatbot dan Pelindungan Hukum terhadap Penggunanya. Jurnal Rechts Vinding: Media Pembinaan Hukum Nasional vol. 12, no. 2, (2023):290-292, doi:http://dx.doi.org/10.33331/rechtsvinding.v12i2.1258.

[38] Washington Simanjuntak dkk., "PERAN PEMERINTAH DALAM IMPLEMENTASI ARTIFICIAL INTELLIGENCE (AI) DI KEMENTERIAN KOMUNIKASI DAN INFORMATIKA REPUBLIK INDONESIA (KEMENKOMINFO RI)," Journal of Social and Economics Research vol. 6, no. 1 (2024): 10-15. https://doi.org/10.54783/jser.v6i1.332.

[39] Bureau of Cyberspace and Digital Policy. Risk Management Profile for Artificial Intelligence and Human Rights. (2024, Juli 25). Diakses dari https://www.state.gov/risk-management-profile-for-ai-and-human-rights/.

[40] Insights. California Enacts AI Transparency Law Requiring Disclosures for AI Content. (2024, Oktober 24). Diakses

Under the Act, Providers (GAI providers) are required to disclose (free of charge) AI detection tools that will allow users to assess whether image, video, or audio content has been created or altered using GAI systems subject to certain technical requirements and privacy protections, these tools will also be required to provide system origin data related to the content, to allow users to verify the device, system, or service used to generate the content and the authenticity of the content. In addition, Providers are also required to disclose latent technical metadata that includes the name of the Provider, the name and version number of the GAI system, the time and date of content creation or change, and a unique identifier. In addition, Providers also include an option for users to explicitly disclose (manifest disclosures) related to information that is clearly, conspicuously, and permanently stated that content was generated by AI.

The first global AI law was presented through the EU AI Act. The Act aims to provide a framework for the development, market placement, and use of AI systems that pose potential risks to health, safety, or human rights. The EU AI Act classifies 4 (four) risk-levels in AI. GAI falls into the limited risk or specific transparency risk category. GAI is classified into general-purpose AI models that can perform a wide variety of tasks and commands.[41]

One of the key aspects of the EU AI Act is the imposition of transparency requirements. Article 53 of the EU AI Act states that AI development companies must fulfill the obligation of transparency of the training data of their AI models, a list of which must be publicly available, or visible on systems intended to interact with users (Article 50 of the EU AI Act). Article 107 underscores the importance of transparency in ensuring accountability and facilitating copyright enforcement. In particular, it mandates AI companies to create and make publicly available a sufficiently detailed summary of the data used to train its AI models. This summary can provide comprehensive information about the data sets used, including both public and private sources, to enable copyright holders to effectively exercise their rights. The template of the summary was developed by the AI Office. The EU AI Act introduced a limited exception for text and data mining on the basis of the importance of balancing copyright protection with encouraging the advancement of innovation and research. However, Article 105 of the EU AI Act underlines the right of copyright holders to prohibit the use of their works for AI training, except for scientific research purposes. Article 105 of the EU AI Act has also explicitly linked the use of works as training data for AI models with the provisions of Article 4 of the Directive on copyright and related rights in the Digital Single Market (DSM Act) which provides for the exclusion of copyright protection for text and data mining purposes.

---

dari https://www.jonesday.com/en/insights/2024/10/california-enacts-ai-transparency-law-requiring-disclosures-for-ai-content#:~:text=The%20Background%3A%20On%20September%2019,or%20altered%22%20using%20generative%20artificial.

[41] David Fernández-Llorca dkk., "An interdisciplinary account of the terminological choices by EU policymakers ahead of the final agreement on the AI Act: AI system, general purpose AI system, foundation model, and generative AI," Artificial Intelligence and Law (2024), https://doi.org/10.1007/s10506-024-09412-y.

In relation to the DSM Act which is the legal basis for copyright protection in the European Union, there is a restriction in Article 3 of the DSM Act on duplication or extraction by research organizations and cultural heritage institutions for scientific research, assuming that the organization has legitimate access to the works and other things intended for data mining purposes, and the copyright owner does not expressly prohibit the use of his work for data mining purposes. However, such exceptions are subject to Article 5 paragraph (5) of Directive 2001/29/EU which states that such limitations or exceptions shall only apply to certain special cases which are not contrary to the legitimate interests of the copyright owner. In Article 15 of the DSM Act, there are new rights granted to EU-based press publishers for the use of digital publications. These rights only apply to use by online service providers and not to personal or non-commercial use by individual users. Inclusion of hyperlinks and very short quotations from press publications are excluded from these rights. Where the use of works included in press publications generates commercial benefits, the authors of works included in press publications shall receive an appropriate share of the revenue derived from their use.

Japan, as one of the countries leading the development of AI, did not create AI-specific regulations. Instead, Japan only revised the relevant laws to address legal issues that have arisen after the development of AI. Regarding the use of copyrighted works as AI training data, Japan revised its Copyright Act (Act No. 48 of May 6, 1970, as amended on January 1, 2022) ("Japanese Copyright Act"). The Japanese Copyright Act does not protect the use of copyrighted works for reporting current events (Article 39), exploitation of political speech (Article 40), and reporting and disclosure of information about ongoing events or activities (Article 41). Meanwhile, AI-related copyright regulations in the Japanese Copyright Act are found in Articles 30-4, 47-4, and 47-5. Article 30-4 of the Japanese Copyright Act states that the use of copyright is not infringing if the exploitation is not to enjoy the expression or allow others to enjoy the expression of the work used. The limitation in this exploitation is that it is prohibited from harming the reasonable interests of the Copyright Owner. Things that are allowed within the limitation include when the Copyrighted Work is used for testing technology related to sound or visual recording; data analysis, such as extraction, comparison, classification, or statistical analysis of language, sound, images, or other data elements; and use in computer data processing or in a way that does not involve the expression of the work that can be captured by human senses (except for the execution of the Copyrighted Work through a computer program).[42]

Section 47-4 permits the use of copyrighted works without the copyright owner's permission to ensure the smooth or efficient use of the copyrighted work on a computer or to maintain or restore its condition of use. Article 47-4 includes specific circumstances as well as a general framework for determining whether unlicensed use is permitted. For

---

[42] Artha Dermawan, "Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese 'nonenjoyment' purposes?," Journal of World Intellectual Property vol. 27, no. 1 (1 Maret 2024): 44–68, https://doi.org/10.1111/jwip.12285.

example, caching to speed up the processing of information over a network and temporary copying of data onto media from a portable audio player when exchanging data to another party may be done without the copyright owner's permission. Article 47-5 permits unlicensed use of copyrighted works where such use is minor and forms part of the computer's processing of information and disclosure of the results. Specifically, Article 47-5 provides that the use must be minor, e.g. when searching for a particular book using specific keywords and displaying a portion of the book with those keywords may be done without the copyright owner's permission.

Thus, within the framework of Japanese copyright law, protection is given to the use of copyrighted works for the purpose of Machine Learning (ML) because such use of copyrighted works is not sensible by human sense/perception, but is limited to the purpose of machine learning. In addition, to accommodate these needs, this exception also covers data scraping or data mining activities carried out through internet transmission, such as access, download, and upload.[43]

By studying the various landscapes of copyright protection in various countries related to the use of Copyrighted Works for GAI training, a common thread can be drawn that can be an alternative regulation for copyright law reform in Indonesia that does not hinder the development of AI. Regulatory arrangements in Indonesia can reflect on regulations in Japan and the European Union that provide relaxation of the use of Copyrighted Works for the benefit of AI development, in the form of relaxation of text and data mining activities. In accommodating the needs of AI development, the use of copyright does not require the author's permission if it includes data scraping or data mining activities carried out through internet transmission, such as access, download, and upload. With this regulation, the act of copying works for the purpose of GAI training as implemented in ChatGPT and Perplexity AI becomes legal. Such relaxation is needed because in training a complex and quality GAI, big data is required, including "vast scrape of most of the internet" or massive and complex data collection from most parts of the internet, which can touch billions of web pages, and this is done by GAI on an automatically programmed algorithm.[44] Thus it is very difficult for GAI to fulfill the element of permission from the Copyright Owner in duplicating data in the form of copyrighted works.

However, regulations must still have restrictive provisions so that AI can be developed without infringing on the reasonable interests of the Copyright Owner. For example, several large sites that have license agreements with copyright owners related to GAI training have included a disclaimer in the license to use the software on the site/application in the End User License Agreement (EULA) before the "user" uses the

---

[43] Wibowo, Richard Jatimulya Alam. Perbandingan Perlindungan Hak Moral dan Hak Ekonomi dalam Pengembangan serta Pemanfaatan Kecerdasan Buatan di Indonesia dengan Jepang, Inggris, Uni Eropa, dan Amerika Serikat. Skripsi. Jakarta: Universitas Tarumanagara. 2024. Hlm : 102.

[44] Tegar Raffi Putra Jumantoro dkk., "Menilik Pro Dan Kontra Pemanfaatan Dan Penetapan Status Hukum Artificial Intelligence (AI) Dalam Hukum Positif Indonesia," Journal of Analytical Research, Statistics and Computation vol. 3, no. 1, (2024): 52. DOI: https://doi.org/10.4590/jarsic.v3i1.28.

software or application so that the burden of responsibility falls on the user of the application or software who has agreed to have their data used. In general, GAI sites have also made a disclaimer on the use of software or applications.

The use of a disclaimer is an early stage in compliance with AI transparency principles if a country is committed to establishing AI in the use of its data systems. AI companies can also be used as legal subjects if there are violations related to intellectual property rights, especially copyright because indirectly the use of disclaimers on some software usage licenses on the site or application (EULA) causes a written agreement between the user of the site or application and the AI company/application provider (tenant) which causes legal consequences between the two legal subjects. AI companies only create tools in the form of GAI that answer all user requests and needs. Therefore, if the GAI company has implemented the preventive measures as mentioned above, the user as the end-user of the tools will be responsible for the use of the requested output.

The inclusion of the disclaimer can be one of the substances for the formulation of the AI Law or similar regulations. In this case, an example can be taken from the substance of the AI Policy Act and the California AI Transparency Act which require GAIs to clearly and explicitly inform users who interact with the GAI regarding its status as a GAI. The disclaimer should explain that there is no plagiarism-free guarantee of the output provided by the GAI chatbot. In addition, for text model GAIs such as Perplexity AI, it should also be made clear that the output and references it provides are not guaranteed to be fully accurate and encourage users to click directly on the links to the answers provided to ensure accuracy. In terms of fulfilling the AI transparency principle, Indonesia can reflect on the EU AI Act that requires AI companies to compile and make available to the public a detailed summary of the data used to train AI models so that copyright holders can enforce their rights more effectively.

However, when referring to copyright law in Indonesia, even though the AI company has fulfilled the principles of transparency, listed the source, and moral rights have been fulfilled, the use of Copyrighted Works must refer to the provisions of fair use in Indonesia, which can be concluded, formulated on 3 (three) conditions, namely 1) not commercial, 2) benefiting the Copyright Owner, 3) the Copyright Owner does not object to the use of their Copyrighted Works. The Copyright Owner does not object to the use of their Copyrighted Works.[45] This is indisputable and should still be championed in the future copyright and AI regulatory landscape to avoid abandonment of the Copyright Owner's rights to their creative work. So in this case, despite the aforementioned preventive measures, if the AI company has not entered into a license agreement with the Copyright Owner of the copyrighted work that has been licensed, and derives economic benefits from the use of the copyrighted work, then the Copyright Owner can still sue the company

---

[45] Riswandi, Budi Agus et al. Pembatasan dan Pengecualian Hak Cipta di Era Digital, Bandung: PT Citra Aditya Bakti, 2017. p.93. https://edeposit.perpusnas.go.id/collection/pembatasan-dan-pengecualian-hak-cipta-di-era-digital-sumber-elektronis/75182#.

to fulfill his economic rights and prove his losses in court.

Simplification of the fulfillment of economic rights can be done by applying a blanket license which is a non-exclusive license for reproduction rights granted by LMK.[46] LMK which was previously regulated in Law 28/2014 is limited to LMK in the field of songs/music. However, the LMK in the field of written works has been regulated in Permenkumham 15/2024 concerning Management of Royalties on Secondary Use Licenses for Copyright of Books and/or Other Written Works. So that the implementation of blanket licenses for written works already has a legal base and can be implemented. With the blanket license, AI developers do not need to enter into a license agreement with each Copyright Owners of the written work, but can pay the royalties collectively to the relevant LMK. The existing legal basis needs to be strengthened by optimizing the LMK in terms of administrative procedures that include data collection of Copyrighted Works, royalty distribution, and license simplification so that the blanket license mechanism runs effectively.

In relation to LMK, it is necessary to harmonize with Law 28/2014 on the unsynchronization between substances, for example regarding overlapping provisions on the use of actual news, where Article 43 of Law 28/2014 allows the use of actual news with the condition of attribution, but Article 18 Permenkumham 15/2024 provides a limit of 10% for the use of actual news, and emphasizes the obligation to pay royalties to LMK if the use is carried out commercially. Synchronization and harmonization between regulations must be done immediately to ensure legal certainty regarding how much portion of the work can be used. The collective payment system through LMK must also be integrated with the POP HC system (Automatic Copyright Registration Approval). This system, which began implementation in 2022, has proven to simplify the process and increase public awareness of the importance of registering their copyrights.

The moral rights of the Copyright Owner must also be fulfilled in the use of the work for GAI training. As in Article 105 of the EU AI Act, it underlines the right of copyright holders to prohibit the use of their works for AI training, except for scientific research purposes. This relates to the transparency mechanism of AI training data, which must be shown in a summary form that makes it easier Copyright Owners and/or Copyright Holders to monitor the use of their work. If a Copyright Owner and/or Holder's work is included in such a summary, and the Copyright Owner and/or Holder objects to or prohibits the use of their copyrighted work, the work should be immediately withdrawn from the relevant AI training dataset. In order to ensure the fulfillment of the Copyright Owner's rights, the implementation of the objection mechanism should be clearly regulated to increase legal certainty.

Overall, apart from the alternative substances that have been offered previously,

---

[46] Ari Juliano Gema, "Masalah Penggunaan Ciptaan Sebagai Data Masukan Dalam Pengembangan Artificial Intelligence Di Indonesia," Technology and Economics Law Journal vol. 1, no. 1 (11 Februari 2022): 12-15. DOI: https://doi.org/10.21143/telj.vol1.no1.1000.

harmonization of laws and regulations must be carried out immediately to keep pace with current developments. The revision of the Copyright Law should also incorporate UNESCO's Readiness Assessment Methodology (RAM) to address the ethical implications of applying AI to a country's database system, particularly in relation to citizens' data. Furthermore, this revision should consider other regulations pertaining to Intellectual Property Rights, such as trademarks, patents, and other related areas, which can be influenced or applied by AI technologies. Indonesia's National Strategy for Artificial Intelligence 2020-2045 by the Agency for the Assessment and Application of Technology (BPPT) can serve as the basis for a more comprehensive law on AI. AI regulations need to classify the level of risk in AI as stipulated in the EU AI Act so that there is a more measurable risk-based approach.

Furthermore, it is crucial to establish an independent mechanism for accuracy verification to assess the accuracy and compliance of AI with copyright regulations. Additionally, it is necessary to synchronize existing regulations and integrate the findings of comparative studies from various countries to inform the reform of the Copyright Law. The relevant regulatory framework must ensure transparency and include supervision mechanisms that involve all stakeholders, including Copyright Owners and the AI industry. This should encompass small independent creators, copyright owners/holders, and MSMEs, particularly digital startups, which require relaxation of web scraping activities to support their operations. In this way, a more responsive and adaptive copyright law framework can be developed, one that aligns with ongoing technological advancements.

CONCLUSION

The validity of using copyrighted works for training GAI in the case study involving lawsuits against OpenAI, Inc. and Perplexity AI, Inc. hinges on the process by which the copyrighted works are inputted into the system (input) and the resulting dataset, which is processed into answers or output responses provided by the GAI to the user. In terms of the input process, under the current copyright law framework in Indonesia, data scraping and/or text and data mining conducted by GAI is considered a form of copying the copyrighted work, as outlined in Article 9 of Law 28/2014, which requires authorization from the copyright owner. Therefore, based on the input process, such actions would be deemed invalid from the perspective of Indonesian copyright law. When examined from the perspective of the output, several elements of infringement may be established depending on the context, including but not limited to: 1) Output that fails to include the source or reference to the relevant copyrighted works violates Article 7 of Law 28/2014, which prohibits the omission of copyright management information; 2) Output that reproduces part or all of a substantial portion of the copyrighted work infringes the economic rights of the copyright owner under Article 9 of Law 28/2014; 3) Output that contains false information regarding the copyrighted work infringes the moral rights of the copyright owner as stipulated in Article 5, as it harms the owner's personal honor or reputation. An alternative that could serve as a middle ground between the

legitimate interests of copyright owners and the needs of AI development is to allow a relaxation in the use of copyrighted works for text and data mining in AI training, while still adhering to the principle of AI transparency as outlined in AI law. AI development companies should also include a disclaimer in their software licenses, presented in the EULA on their websites or applications. A balance between the economic interests of copyright owners and/or holders and the advancement of GAI can be achieved through the enforcement of a blanket license via the LMK, as regulated in Permenkumham 15/2024, which should be integrated with the POP HC system. These regulations, together with other existing legal frameworks, can be synchronized and/or integrated, alongside reforms derived from comparative studies of international regulations, to update the Copyright Law in a manner that is more adaptive to technological advancements.

ACKNOWLEDGEMENT

1. Christine S.T. Kansil, S.H., M.Hum., who has taken the time and contributed in pouring out thoughts and energy in guiding the author so that this article can be completed properly;
2. The resource persons who have taken the time to conduct interviews with the author so that they have contributed in expressing ideas and answers related to the problems studied in this research;
3. The authors' families, who provided moral, spiritual and material support until this article could be completed.

REFERENSI

Amerika Serikat. Title 17 of the United States Code (Digital Management Copyright Act). https://www.copyright.gov/title17/.

Alfani, Ninda et al. Implementasi UU Nomor 28 Tahun 2014 dalam Perlindungan Hak Cipta di Era Digital. *Journal of Administrative and Social Science*, vol. 4 no.1 (2022): 23-36. DOI: https://doi.org/10.55606/jass.v4i1.

Amazon. What is Natural Language Processing (NLP)?. Diakses dari https://aws.amazon.com/what-is/nlp/.

Andreotta, Adam J., Nin Kirkham, dan Marco Rizzi. "AI, big data, and the future of consent." *AI and Society* vol. 37, no. 4 (1 Desember 2022): 1715–28. https://doi.org/10.1007/s00146-021-01262-5.

Amerika Serikat. *Surat Gugatan Dow Jones & Company, Inc.; NYP Holdings, Inc., vs Perplexity AI, Inc.*

Amerika Serikat. *Surat Gugatan OpenAI v. Raw Story Media, Inc dan AlterNet Media, Inc.*

Bureau of Cyberspace and Digital Policy. Risk Management Profile for Artificial Intelligence and Human Rights. (2024, Juli 25). Diakses dari https://www.state.gov/risk-management-profile-for-ai-and-human-rights/.

Bureau of Cyberspace and Digital Policy. Risk Management Profile for Artificial Intelligence and Human Rights. (2024, Juli 25). Diakses dari https://www.state.gov/risk-management-profile-for-ai-and-human-rights/.

California, Amerika Serikat. Bill SB-942: California AI Transparency Act. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942.

Camilleri, Mark Anthony. "Artificial intelligence governance: Ethical considerations and implications for social responsibility." *Expert Systems* vol. 41, no. 7 (1 Juli 2024). https://doi.org/10.1111/exsy.13406.

Coeckelbergh, Mark. "Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability." *Science and Engineering Ethics* vol. 26, no. 4 (1 Agustus 2020): 2051–2068. https://doi.org/10.1007/s11948-019-00146-8.

Dermawan, Artha. "Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese 'nonenjoyment' purposes?" *Journal of World Intellectual Property* vol. 27, no. 1 (1 Maret 2024): 44–68. https://doi.org/10.1111/jwip.12285.

Devina, Raden Radisa Difa & Ramli, Tatty Aryani. "Penggandaan dan Pengumuman Karya Cipta *E-Book* berdasarkan Prinsip *Fair Use* Ditinjau dari Hukum Positif." *Bandung Conference Series: Law Studies*, vol. 2 no. 1, (2022): 457-464. DOI: https://doi.org/10.29313/bcsls.v2i1.890.

Djulaeka & Rahayu, Devi. Buku Ajar: Metode Penelitian Hukum. Surabaya : Scopindo Media Pustaka, 2020. pp. 88. *https://books.google.co.id/books?id=aIrUDwAAQBAJ&printsec=copyright&redir_esc=y#v=onepage&q&f=false*.

Emilia David. The Intercept, Raw Story, and AlterNet sue OpenAI and Microsoft. Feb 29, 2024. https://www.theverge.com/2024/2/28/24085973/intercept-raw-story-alternet-openai-lawsuit-copyright.

Fernández-Llorca, David, Emilia Gómez, Ignacio Sánchez, dan Gabriele Mazzini. "An interdisciplinary account of the terminological choices by EU policymakers ahead of the final agreement on the AI Act: AI system, general purpose AI system, foundation model, and generative AI." *Artificial Intelligence and Law*. Springer Nature, 2024. https://doi.org/10.1007/s10506-024-09412-y.

Fitria, Tira Nur. "Using ChatBot-Based Artificial Intelligence (AI) for Writing an English Essay: The Ability of ChatGPT, Perplexity AI, and ChatSonic." *Journal of Language Intelligence and Culture* vol. 6, no. 2 (26 Agustus 2024): 103–128. https://doi.org/10.35719/jlic.v6i2.139.

François Fleuret. The Little Book of Deep Learning. Universite de Geneve, 2024. p. 12.

Gema, Ari Juliano. "Masalah Penggunaan Ciptaan Sebagai Data Masukan Dalam Pengembangan Artificial Intelligence Di Indonesia." *Jurnal Hukum & Pembangunan* vol. 1, no. 1 (11 Februari 2022). https://doi.org/10.21143/telj.vol1.no1.1000.

Hikmah, Faidatul, et al. Perlindungan Hak Ekonomi Bagi Pemilik Hak Cipta Dalam Perspektif Hukum Kekayaan Intelektual Di Indonesia. Jurnal Pendidikan Dan Konseling

(JPDK) vol. 5 no. 2, (2023): 2254–2260. DOI: https://doi.org/10.31004/jpdk.v5i2.13503.

HukumOnline. Kemenangan bagi Pemegang Karya Tulis dengan Hak Cipta! Batasan Penggunaan Lisensi Sekunder Ditetapkan 10%. (2024, Juli 5). Diakses dari https://pro.hukumonline.com/a/lt6687682cc3d63/kemenangan-bagi-pemegang-karya-tulis-dengan-hak-cipta-batasan-penggunaan-lisensi-sekunder-ditetapkan-10.

Hyden, Kan. "AI, Norms, Big Data, and the Law." *Asian Journal of Law and Society* 7, no. 3 (1 November 2020): 409–36. https://doi.org/10.1017/als.2020.36.

Indonesia, Undang-undang Nomor 28 Tahun 2014 tentang Hak Cipta (Lembaran Negara Republik Indonesia Tahun 2014 Nomor 266, Tambahan Lembaran Negara Republik Indonesia Nomor 5599). https://peraturan.bpk.go.id/Details/38690.

Indonesia, Peraturan Menteri Hukum dan HAM Nomor 15 Tahun 2024 tentang Pengelolaan Royalti atas Lisensi Penggunaan Sekunder untuk Hak Cipta Buku dan/atau Karya Tulis Lainnya.

Indonesia, Undang-undang Nomor 28 Tahun 2014 tentang Hak Cipta (Lembaran Negara Republik Indonesia Tahun 2014 Nomor 266, Tambahan Lembaran Negara Republik Indonesia Nomor 5599).

Insights. California Enacts AI Transparency Law Requiring Disclosures for AI Content. (2024, Oktober 24). Diakses dari https://www.jonesday.com/en/insights/2024/10/california-enacts-ai-transparency-law-requiring-disclosures-for-ai-content#:~:text=The%20Background%3A%20On%20September%2019,or%20altered%22%20using%20generative%20artificial.

Iorliam, Aamo & Joseph Abunimye Ingio. "A Comparative Analysis of Generative Artificial Intelligence Tools for Natural Language Processing." *Journal of Computing Theories and Applications* 1, no. 3 (26 Februari 2024): 311–325. https://doi.org/10.62411/jcta.9447.

Jamaaluddin, MM & Sulistyowati, Indah. Buku Ajar Kecerdasan Buatan (Artificial Intelligence). Sidoarjo : Umsida Press. (2021).

Jepang. Copyright Act (Act No, 48 of 1970) as Amended in 2022. https://www.wipo.int/wipolex/en/legislation/details/21342.

Jumantoro, Tegar Raffi Putra. "Menilik Pro Dan Kontra Pemanfaatan Dan Penetapan Status Hukum Artificial Intelligence (AI) Dalam Hukum Positif Indonesia". *Journal of Analytical Research, Statistics and Computation* vol. 3 no. 1, (2024): 51-70. DOI: https://doi.org/10.4590/jarsic.v3i1.28.

Juliardi, Budi et al. Metode Penelitian Hukum. Sumatera Barat : CV. Gita Lentera, 2023. p. 43. https://books.google.co.id/books?id=vyXbEAAAQBAJ&printsec=frontcover&hl=id#v=onepage&q&f=false.

Kahveci, Zeynep Ülkü. Attribution problem of generative AI: a view from US copyright law.

Journal of Intellectual Property Law and Practice 18, no. 11 (1 November 2023): 796–807. https://doi.org/10.1093/jiplp/jpad076.

Lucchi, Nicola. ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems. European Journal of Risk Regulation, (2023). Hlm : 1–23. DOI: 10.1017/err.2023.59.

Longoni, Chiara, Andrey Fradkin, Luca Cian, dan Gordon Pennycook. "News from Generative Artificial Intelligence Is Believed Less." Dalam *ACM International Conference Proceeding Series*, 97–106. Association for Computing Machinery, 2022. https://doi.org/10.1145/3531146.3533077.

Nainggolan, Bernard. *Perlindungan Hukum Kekayaan Intelektual di Era Digital*. Yogyakarta : Publika Global Media. (2021).

Putra, G., Taniady, V., & Halmadiningrat, I. "Tantangan Hukum: Keakuratan Informasi Layanan AI Chatbot dan Perlindungan Hukum terhadap Penggunanya". *Jurnal Rechts Vinding: Media Pembinaan Hukum Nasional*, vol. 12 no. 2, (2023). DOI: http://dx.doi.org/10.33331/rechtsvinding.v12i2.1258.

Rizkia, Nanda Dwi & Fardiansyah, Hardi. *Hak Kekayaan Intelektual Suatu Pengantar*. Bandung: Penerbit Widina. (2022).

Rahayu, Kanti. "Kajian Hukum Tentang Penerapan Fungsi Sosial Pada Hak Cipta." *Diktum: Jurnal Ilmu Hukum* vol. 9 no. 2 (22 Agustus 2022): 150–167. https://doi.org/10.24905/diktum.v9i2.92.

Riswandi, Budi Agus et al. Pembatasan dan Pengecualian Hak Cipta di Era Digital. Bandung: PT Citra Aditya Bakti. (2017). https://edeposit.perpusnas.go.id/collection/pembatasan-dan-pengecualian-hak-cipta-di-era-digital-sumber-elektronis/75182#.

Shtykalo, O & Yamnenko, I. "ChatGPT and Other AI Tools for Academic Research and Education." *Digital Ecosystems: Interconnecting Advanced Networks with AI Applications*, Vol. 1198, (2024). DOI: https://doi.org/10.1007/978-3-031-61221-3_29.

Shukla, Manali, Ishika Goyal, Bhavya Gupta, dan Jhanvi Sharma. "A Comparative Study of ChatGPT, Gemini, and Perplexity." *International Journal of Innovative Research in Computer Science and Technology* vol. 12, no. 4 (Juli 2024): 10–15. https://doi.org/10.55524/ijircst.2024.12.4.2.

Simanjuntak, Washington, Agus Subagyo, & Dadang Sufianto. "PERAN PEMERINTAH DALAM IMPLEMENTASI ARTIFICIAL INTELLIGENCE (AI) DI KEMENTERIAN KOMUNIKASI DAN INFORMATIKA REPUBLIK INDONESIA (KEMENKOMINFO RI)". Journal of Social and Economics Research vol. 6 no. 1, (2024): 1-15. https://doi.org/10.54783/jser.v6i1.332.

Tyagi, Kalpana. "Copyright, text & data mining and the innovation dimension of generative AI." *Journal of Intellectual Property Law and Practice* vol. 19, no. 7 (1 Juli 2024): 557–70. https://doi.org/10.1093/jiplp/jpae028.

Taecharungroj, Viriya. "'What Can ChatGPT Do?' Analyzing Early Reactions to the Innovative AI Chatbot on Twitter." *Big Data and Cognitive Computing* vol. 7, no. 1 (1 Maret 2023). https://doi.org/10.3390/bdcc7010035.

Triatmojo, Firmandanu et al. Perlindungan Hak Cipta Lagu Komersil. Jawa Tengah : NEM. (2021). P. 9.

European Union. Directive (EU) 2019/790 on Copyright in The Digital Single Market. https://eur-lex.europa.eu/eli/dir/2019/790/oj.

European Union. AI Act of 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

Utah, Amerika Serikat. AI Policy Act of 2024. https://le.utah.gov/xcode/Title13/Chapter72/C13-72_2024050120240501.pdf

Wijayanti, Punik Triesti & Kharisma, Dona Budi. "Web Scraping dalam Aplikasi ChatGPT oleh Chatbot Berbasis Artificial Intelligence (AI) Berdasarkan Undang-Undang Nomor 28 Tahun 2014 Tentang Hak Cipta." *Souvereignty : Jurnal Demokrasi dan Ketahanan Nasional,* vol. 3 no. 2, (2024): 114-121. DOI: . https://doi.org/10.20961/souvereignty.v3i2.1620.

Wibowo, Richard Jatimulya Alam. Perbandingan Perlindungan Hak Moral dan Hak Ekonomi dalam Pengembangan serta Pemanfaatan Kecerdasan Buatan di Indonesia dengan Jepang, Inggris, Uni Eropa, dan Amerika Serikat. Skripsi. Jakarta: Universitas Tarumanagara. 2024.

Zhang, Caiming & Lu, Yang. Study on Artificial Intelligence: The State of The Art and Future Prospects. Journal of Industrial Information Integration, Vol. 23, Artikel 100224. https://doi.org/10.1016/j.jii.2021.100224.