

**COPYRIGHT IN THE AGE OF AI:
EXAMINING LIABILITIES OF AI ACTORS IN THE
UNAUTHORIZED USES OF COPYRIGHTED WORKS FOR
MODEL TRAINING THROUGH THE LENS OF
LIBERTY AND PROSPERITY**

CHAIRA FAYE L. FERRAN

A thesis presented in partial fulfillment
of the requirements for the degree of

JURIS DOCTOR



DE LA SALLE LIPA
College of Law

FOUNDATION FOR LIBERTY AND PROSPERITY

2024 - 2025 DISSERTATION WRITING CONTEST

ABSTRACT

Recent developments in artificial intelligence (AI) have introduced significant copyright concerns, particularly on the liabilities of AI actors involved in unauthorized uses of copyrighted materials as training datasets for AI models. Since AI systems depend heavily on large-scale data for effective machine learning, it is crucial to explore the intersection of AI technologies and copyright law in order to develop adaptive legal frameworks that respond to the evolving landscape of creative expression and knowledge production. This study delves into four critical areas: global AI legislation, potential copyright law exemptions in AI, the responsibilities of AI actors in the unauthorized uses of copyrighted materials, and amendments in the Intellectual Property Code of the Philippines (IPC).

This research employed qualitative and black letter approach to establish a regulatory framework governing AI actors' liability for unauthorized uses of copyrighted content. Both primary and secondary authorities were utilized, including the IPC, international copyright laws, books, legal journals, judicial decisions, AI frameworks, and expert analyses. Interviews with the director of copyright, legal scholars and practitioners specializing in copyright and AI models were likewise utilized to provide further insights.

The findings indicate substantial disparities in international approaches to AI regulation, with some jurisdictions enacting AI-specific provisions related to copyright while others continue to apply traditional copyright laws. Although certain legal frameworks allow the use of copyrighted materials for AI training under specific conditions, they often fall short of comprehensively addressing the liability of the AI actors. Nonetheless, it revealed that key responsibilities of AI developers include promoting transparency and data disclosure, ensuring compliance with copyright laws, maintaining data quality, and upholding accountability measures.

The study also highlights the dual responsibilities of AI actors during both the training (input) and utilization (output) stages, with

primary responsibility for unauthorized uses lying with AI developers and development companies, including data collectors and model trainers. Users may also face direct liability for infringing acts. Various permissible acts under the international AI legislations serve as exemptions to general copyright requirements, such as transparency for general purpose AI models, text and data mining, use of data for non-professional and scientific research, use of data for non-enjoyment, computer data processing, incidental use of data for efficient operation, and training data for creating new knowledge.

Identified exceptions to permissible acts include reserving rights, ensuring fair competition, attribution to copyright holders, avoiding reasonable harm, and considering purpose and interference. Effective enforcement of these rights likewise requires robust mechanisms and regular audits, with recommended measures such as valuation of copyright holders' work, disclosure requirements, dataset registration with the Intellectual Property Office (IPO), fair learning options, and digital receipts via blockchain.

In order to address gaps, the study recommends integrating machine learning provisions into the IPC, creating Implementing Rules and Regulations (IRR) to empower IPO to enforce disclosure requirements and registration of dataset, and embedding digital rights signatures in copyrighted materials. These measures aim to balance innovation with protecting creators' rights, fostering a legally compliant and ethical AI ecosystem that respects intellectual property in the digital age.

Keywords: *Artificial intelligence, copyright law, intellectual property, AI legislation, machine learning, training data, fair use*

TABLE OF CONTENTS

ABSTRACT	1
OUTLINE.....	5
INTRODUCTION.....	8
STATEMENT OF THE PROBLEM.....	10
OBJECTIVES OF THE STUDY	11
SIGNIFICANCE OF THE STUDY.....	12
SCOPE AND LIMITATIONS.....	13
METHODOLOGY	13
CHAPTER I: ARTIFICIAL INTELLIGENCE AND COPYRIGHT	
LAW	15
Copyright Law	16
Copyright Infringement.....	17
Understanding copyright infringement in the use of training data	20
Defining copyright infringement in the use of training data	21
CHAPTER II: AI LEGISLATION AROUND THE GLOBE: LIABILITIES OF AI ACTORS IN UNAUTHORIZED USES OF COPYRIGHTED MATERIALS	24
United States.....	24
Europe	25
United Kingdom	26
Japan	27
China.....	28
Singapore	29
Australia	30
Philippines	31
CHAPTER III: LEGAL RAMIFICATIONS OF COPYRIGHTED MATERIALS IN AI: POSSIBLE LEGAL EXEMPTIONS.....	47
Text and Data Mining	48

Use of copyrighted data for non-professional and scientific research.....	50
Use of work for non-personal enjoyment.....	51
Use of work in computer data processing	52
Use of data for incidental purposes to ensure smooth and efficient operation	53
Use of materials for training data to create new knowledge through computerized data processing.....	54
Use of copyrighted data for model training with different purposes	54
CHAPTER IV: BETWEEN CODE AND CREATIVITY: LEGAL RESPONSIBILITIES OF AI ACTORS FOR THE UNAUTHORIZED USES OF COPYRIGHTED MATERIALS	60
Copyright Infringement on the Different Stages of AI process	61
The Responsibilities of AI Actors in the Unauthorized Uses of Copyrighted Materials	66
CHAPTER V: SAFEGUARDING INNOVATION: UPDATING THE PHILIPPINE IP CODE FOR THE AGE OF AI.....	74
CHAPTER VI: SUMMARY, CONCLUSION, AND RECOMMENDATIONS.....	83
Balancing Copyright and AI: Navigating Liability in the Digital Age	83
Copy, Right? Proposals for Harmonizing Copyright and AI.....	85
ANNEX: PROPOSED RECOMMENDATIONS TO THE AMENDMENT OF REPUBLIC ACT NO. 8293.....	88
BIBLIOGRAPHY.....	99

OUTLINE

This thesis focuses on the *“input”* aspect of AI-generated works, specifically, the use of copyrighted materials as training data for AI models. It espouses the idea of balance between liberty and prosperity to promote innovation through responsible AI use while protecting the integrity of original creative works.

CHAPTER I

This chapter explores the intersection of artificial intelligence (AI) and copyright law, focusing on the challenges that arise from the use of copyrighted materials in AI development. It begins with a general overview of copyright principles and the concept of copyright infringement, particularly in the context of training data used by AI systems.

The chapter also discusses how potential infringement can occur at two key stages of AI: (1) during the input phase, where copyrighted data may be reproduced for training purposes, and (2) in the output phase, where AI-generated content may resemble or adapt existing copyrighted works.

CHAPTER II

This chapter analyzes international legal frameworks concerning the liability of AI actors for the unauthorized uses of copyrighted materials in AI models. It evaluates current international AI frameworks and legislations to determine how different jurisdictions address copyright related concerns.

Through a comparative analysis of selected AI legislations, the chapter reveals that some countries have enacted specific provisions to address copyright issues related to AI, while others continue to rely on their existing copyright laws. It also notes that most frameworks and AI

laws do not extensively address the liabilities of AI actors, as they often allow or encourage the use of copyrighted materials in training datasets – explicitly or implicitly – under specific conditions.

CHAPTER III

Building on the findings of the comparative analysis in the previous section, this chapter explores the possible exemptions in AI legislations that permit the use of copyrighted materials in datasets for producing AI-generated works. The researcher discussed the identified permissible acts alongside the required conditions and exceptions, as supported by the narratives of the participants, jurisprudences, and other related studies.

The findings show that the conditions to the permissible acts function as exemptions in the use of copyrighted materials for data training, similar as to how fair use doctrine is utilized in the Philippine jurisdiction.

CHAPTER IV

This chapter investigates the responsibilities of AI actors for damages resulting from the unauthorized uses of copyrighted materials in AI systems. To structure the analysis, the researcher adopted the *Generative AI Supply Chain* as a framework in discussing copyright infringement on the different stages of AI process.

Drawing on jurisprudence, participant interviews, and related studies, the chapter identifies the distinct responsibilities of various AI actors – such as data collectors, developers, and model trainers – across the different stages of the AI development process.

The findings show that responsibilities differ between the training and utilization phases, with primary accountability falling on developers, data collectors, and model trainers during the training stage. Non-

compliance with these responsibilities may likewise lead to copyright infringement.

CHAPTER V

This chapter discusses the proposed amendments to the Intellectual Property Code (IPC) to address the emerging legal challenges surrounding the use of copyrighted materials in AI development. The proposals are grounded in a comprehensive review of primary and secondary legal sources as well as the enriched insights from the interviews with the Director of Copyright and Related Rights, legal scholars, and practitioners with expertise in copyright and AI-related intellectual property law.

It underscores the urgent need to strengthen the protection of copyright holders by clearly defining the boundaries of permissible uses and the conditions under which they apply. The proposed amendments are focused on the three fundamental aspects of copyright – *credit, compensation, and consent*.

CHAPTER VI

This chapter presents key policy recommendations aimed at strengthening the legal framework for copyright protection in the age of artificial intelligence. These recommendations include: (1) integrating specific provisions on machine learning into the IPC, (2) developing comprehensive Implementing Rules and Regulations (IRR) to enhance the authority of the Intellectual Property Office (IPO) in exercising its administrative and adjudicatory functions, particularly in managing repository of copyrighted materials, and (3) embedding a digital rights signature layer in copyrighted works to trace their provenance and enable proper attribution.

These proposals aim to ensure more adaptive and responsive copyright regime in the context of rapidly evolving AI technologies.

INTRODUCTION

The emergence of new technologies and continuous advancements in artificial intelligence (AI) has brought forth several copyright concerns that needed attention.¹ One notable discourse on this topic includes the intricate question of authorship relating to the works produced by the AI systems.² Oftentimes, this raises questions about the legal recognition of the AI as a creative entity and issues on copyright ownership and protection.³

On the opposite side of this spectrum is the input aspect of AI-generated outputs that must likewise be given critical consideration. AI systems are typically trained on vast datasets that include copyrighted materials like texts, artworks, photographs, videos, and audios, often scraped from the internet without permission or licensing.⁴ The unauthorized reproduction and use of these works in the training phase of machine learning may constitute direct or contributory infringement, especially when the AI's outputs replicate or are derived from protected content. The boundary between fair use and infringement remains blurred, hence, understanding the rights and permissions associated with the input has become an essential discourse. In this dynamic landscape of technological innovation and developments, an examination of the interplay between the AI technologies and copyright laws has become indispensable for

-
- ¹ See Jae-Hyun Kim. *A Study on the Copyright issues related to AI (Artificial Intelligence) creation in the era of the 4th industrial revolution*. Journal of Next-generation Convergence Information Services Technology (2021). <https://doi.org/10.29056/jncist.2021.08.06>. See also Rita Matulionyte. *Australian Copyright Law Impedes the Development of Artificial Intelligence: What Are the Options?* IIC - International Review of Intellectual Property and Competition Law, 52 (2020): 417 - 443. <https://doi.org/10.1007/s40319-021-01039-9>.
 - ² Y. Bysaga, D. Byelov and V. Zaborovskyi. *Artificial intelligence and copyright and related rights*. Uzhhorod National University Herald. Series: Law (2023). <https://doi.org/10.24144/2307-3322.2022.76.2.47>.
 - ³ E. Vavilin. *The status of artificial intelligence: from object to the subject of legal relations*. Vestnik Tomskogo Gosudarstvennogo Universiteta. Pravo (2022). <https://doi.org/10.17223/22253513/45/10>.
 - ⁴ KIM, *supra* note 1. See also Enrico Bonadio and Luke McDonagh. *Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity*. ERPN: Intellectual Property (2020).

developing legal frameworks that can adapt to the evolving nature of creative expression and knowledge production.⁵

In the Philippine legal landscape, these challenges are compounded by the absence of explicit statutory provisions regulating AI-generated content or the permissible scope of data use in machine learning.⁶ This legal vacuum exposes not only gaps in the protection afforded to content creators but also creates uncertainties for developers and enterprises relying on AI innovation. Without clear guidelines, creators may find it difficult to assert their rights, while AI developers may inadvertently expose themselves to liability.

This complex dynamic calls for a legal framework that balances innovation with rights protection. It is within this context that the philosophy of the Foundation for Liberty and Prosperity (FLP), championed by retired Chief Justice Artemio V. Panganiban becomes especially relevant.⁷ In *Unleashing Entrepreneurial Ingenuity*, Justice Panganiban emphasizes the importance of crafting laws and policies that simultaneously uphold individual liberties and promote national prosperity.⁸ His vision encourages a regulatory approach that does not stifle technological advancement but ensures it operates within an ethical and rights-based framework. Applying this philosophy to AI and copyright means crafting legislation that both foster innovation and shields creative works from misuse.

⁵ Bohan Shi and Nazura Abdul Manap. *Research On Copyright Problem of Artificial Intelligence Generations in China*. *Journal of Information System and Technology Management* (2023). <https://doi.org/10.35631/jistm.831003>.

⁶ Patrizia Anne Garcia. "Liberty in Creativity and Innovation, Prosperity in Protection Under the Rule of Law." *Thesis*, De La Salle Lipa, 2022. See also Rachel Lois Gella. "Defining the Legal Role and Liability of Artificial Intelligence in Digital Consent towards Liberty and Prosperity in the Philippines." *Thesis*, University of St. La Salle, 2024.

⁷ Artemio V. Panganiban, *Unleashing Entrepreneurial Ingenuity*, 12th General Assembly of the ASEAN Law Association, Makati Shangri-la Hotel, Makati City, February 26, 2015, available at <https://cjpanganiban.com/2015/02/26/unleashing-entrepreneurial-ingenuity/> [last accessed 11 May 2025].

⁸ PANGANIBAN, *supra* note 7.

STATEMENT OF THE PROBLEM

The lack of statute and definite guidelines in regulating the use of artificial intelligence made it difficult to navigate copyright issues in the Philippines. The lack of specific provisions governing this matter likewise threatens the protection accorded to the copyrighted materials. Furthermore, it exposes potential pitfalls and ambiguities in the interpretation of the intellectual property law and related legislation.

To aid in the regulation of AI use, this research aims to bridge these gaps through an examination of both the international and domestic laws concerning the liability of the AI actors in the unauthorized uses of copyrighted materials as data sets for training AI models. This study also seeks to propose amendment in the IPC concerning the use of copyrighted materials as training data for machine learning purposes. Specifically, it seeks to address the following queries:

1. What are the liabilities of AI actors in the unauthorized uses of copyrighted materials to AI models under the:
 - a. Philippine laws
 - b. International laws
2. What are the possible exemptions in the copyright law concerning the unauthorized uses of copyrighted materials to AI models?
3. Which among the AI actors should bear the responsibility for the damages resulting from the unauthorized uses of copyrighted materials to AI models?
4. What improvements can be incorporated to the Intellectual Property Code of the Philippines to cater the challenges brought by the artificial intelligence?

In order to contribute insights in this prevailing contemporary issue, the researcher endeavored to understand artificial intelligence in the context of copyright law, particularly in the use of copyrighted materials in producing AI-generated works.

OBJECTIVES OF THE STUDY

This study attempted to propose amendments in the IPC concerning the liability of the AI actors in the unauthorized uses of copyrighted materials to produce AI-generated outputs. Specifically, this research delved into four areas.

First, it seeks to examine the international legal frameworks governing the liability of AI actors in the unauthorized uses of copyrighted materials to AI models. This was accomplished by scrutinizing the prevailing international AI regulations and other legal approaches.

Second, it aims to evaluate the possible exemptions associated with incorporating copyrighted materials into the datasets used for creating AI-generated works. This was completed through examination of existing international laws, non-binding instruments, foreign judicial decisions, writings of experts, and interviews with the proper authorities.

Third, it intends to determine the responsibility of the AI actors for the damages resulting from the unauthorized uses of copyrighted materials to AI models. This was done through analyses of the international AI regulations, the domestic laws concerning copyright, foreign judicial decisions, writings of legal scholars, and interviews with the proper authorities.

Fourth, it strives to propose amendments to the IPC. This was undertaken using the primary and secondary authorities as well as interviews with the legal scholars and practitioners who have

expertise on copyright and the use of artificial intelligence in intellectual property law.

This thesis espouses the idea of balance between innovation and copyright protection. It promotes the development of policies that enable responsible AI use and protect the integrity of original creative works in a rapidly evolving digital environment.

SIGNIFICANCE OF THE STUDY

The creation of a regulatory framework governing the liability of AI actors in the unauthorized uses of copyrighted materials as datasets to produce AI-generated outputs may mitigate the misuse of copyrighted materials and reduce other risks associated with it while promoting a more secured economy that fosters production of original creative works.

This study is beneficial to different sectors of the society because it can serve as a valuable reference for amending or enhancing the current copyright laws in the Philippines, particularly the Republic Act 8293 or the Intellectual Property Code of the Philippines (IPC).

In light of dynamic advancements in technology and the capricious nature of output production and knowledge generation, this research can be a basis for the creation of copyright guidelines to AI use in order to safeguard individuals against misappropriation of their creative and artistic works. Additionally, the study has the potential to establish legal precedents that can be used as a foundation for addressing liability issues associated with AI use and copyright infringement.

The findings of the research can likewise assist the legislators in crafting informed policies and regulations governing the use of copyrighted materials in AI models. It may influence the creation of legal frameworks that strike the balance between innovation and copyright protection.

SCOPE AND LIMITATIONS

This research is limited by the fact that the area of study is relatively complex and novel. The liability associated with the unauthorized uses of copyrighted materials in the production of AI-generated work is not widely discussed in the literature. Although there are cases of copyright infringement, there is no specific instance when the current copyright law of the Philippines is applied to infringing AI technologies. There is also no individual case mentioning that liability for copyright infringement should be on the person who creates, owns, develops, or uses AI-systems when copyrighted materials are used as datasets to produce AI-generated outputs.

This study is tied entirely on the possible risk of copyright infringement in the use of copyrighted materials as datasets to produce AI-generated works. It is likewise limited to the use of existing international frameworks, copyright infringement cases, and related judicial decisions to determine the liabilities of the AI actors.

METHODOLOGY

Qualitative research design and a black letter approach were employed in an attempt to create a regulatory framework governing the liability of the AI actors in the unauthorized uses of copyrighted materials to produce AI-generated outputs.

Qualitative research is an approach that is used for exploring the meaning and understanding the meaning of individuals or groups ascribe to a social or human problem.⁹ Conversely, a black letter approach to legal research is a traditional and doctrinal method that concentrates on the strict interpretation and analysis of legal texts.¹⁰ It

⁹ John Creswell and David Creswell. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 5th ed. SAGE Publications, 2017.

¹⁰ Terry Hutchinson. (2010). *Researching and writing in law* (3rd ed.). Thomson Reuters Lawbooks, Australia.

involves organizing legal principles into coherent categories and interpreting them based on authoritative legal sources, making it especially useful for developing legal arguments and understanding the structure of the legal systems.¹¹ The exploration of the topic through these approaches enabled the researcher to make comparisons of the similarities and differences of the existing international laws on copyright and AI use vis-à-vis the domestic laws of the Philippines on the protection of intellectual property.

The governing international laws and statutes containing provisions on AI use in copyrighted works were examined through the doctrinal legal research method. Related provisions were list down and compared with each other. Similar provisions were grouped together and compared with the copyright laws of the Philippines. The similarities and differences between and among each other were analyzed according to the research questions of the study.

Interviews among legal scholars and practitioners who have expertise on copyright and the use of artificial intelligence were also conducted. The interviews include narratives from the Director of the Bureau of Copyright and Related Rights of Intellectual Property Office of the Philippines (IPOP HL), a private law practitioner specializing on copyright cases, and an international lawyer on copyright law and artificial intelligence.

The researcher employed thematic analysis in making sense of the narratives of the participants. The data gathered from the interviews were used to support or contradict the result of the doctrinal method. This method of data analysis gave the researcher an idea on how copyright laws are applied in AI use both theoretically and practically.

¹¹ HUTCHINSON, *supra* note 10.

CHAPTER I

ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW

With the proliferation of various AI technologies, artificial intelligence has also emerged as one of the widely discussed subject matter across various legal domains.¹² In copyright law, it presents two significant considerations, particularly on the copyright protection of AI-generated outputs and on the creation of works with the help of AI.¹³

First, there is a discourse as to the protection accorded by the copyright laws to AI-generated outputs.¹⁴ On one hand, some scholars argue that AI works should not be protected under the copyright law since it is designed to safeguard human creations.¹⁵ On the other hand, there are claims that AI works should be afforded with similar protection in order to conserve investments in related industries and to promote industrial development.¹⁶

Second, there is also an apprehension of copyright infringement that occurs on the different stages of data mining used for the development of artificial intelligence.¹⁷ It has been highlighted that the absence of statute governing AI use creates uncertainty as to the exemptions of the said data against fair use provisions of the international laws and the Intellectual Property Code.¹⁸

Given the expansive nature of the subject, this research opts to focus on the second discourse, particularly on the utilization of copyrighted materials as data inputs in generating AI outputs.

¹² Spindler, G. *Copyright Law and Artificial Intelligence*. IIC - International Review of Intellectual Property and Competition Law, 50 (2019): 1049 - 1051.
<https://doi.org/10.1007/s40319-019-00879-w>

¹³ SPINDLER, *supra* note 12.

¹⁴ KIM, *supra* note 1.

¹⁵ Virendra Ahuja. *Artificial intelligence and copyright: Issues and challenges*. ILI Law Review, Winter Issue 2020, <https://ssrn.com/abstract=3864922>

¹⁶ KIM, *supra* note 1.

¹⁷ *Id.*

¹⁸ *Id.*

Copyright Law

Copyright is a form of intellectual property that grants exclusive rights to creators over their literary, artistic, musical, and other creative works.¹⁹ It grants creators a bundle of exclusive legal rights, including the right to reproduce, distribute, perform, and display their works publicly.²⁰ These rights are fundamental to the protection and commercial exploitation of creative works.²¹

Copyright also encompasses both economic rights and moral rights. Economic rights allow creators to benefit financially while moral rights protect personal and reputational interests of the creator.²² The former can be transferred while the latter typically remains with the creator.²³

The primary purpose of copyright is to encourage the creation of new works by granting creators exclusive rights to their creations, thus providing them with economic incentives.²⁴ International treaties, such as the Berne Convention,²⁵ play a crucial role in harmonizing copyright laws across different jurisdictions as it establishes minimum standards of protection and facilitate the international enforcement of copyright.²⁶ The World Intellectual Property Organization²⁷ (WIPO), on the other hand, oversees many of these international agreements.

¹⁹ “Copyright.” *NBER International Seminar on Macroeconomics* 8, no. 1 (2011): iv–iv. <https://doi.org/10.1086/667618>.

²⁰ Hemant Arya and Tarun Kumar Bhatt. *Introduction of intellectual property rights*. (2021). <https://10.1016/B978-0-12-821471-8.00022-2>

²¹ ARYA & BHATT, *supra* note 20.

²² Safet Emruli, Agim Nuhiu and B. Kadriu. *Copyright and Copyright Protection*. *European Journal of Interdisciplinary Studies*, 6 (2016): 36–40.

²³ *Id.* at 38.

²⁴ L. Ray Patterson and Stanley W. Lindberg. *The Nature of Copyright: A Law of Users' Rights*. The Michigan Law Review Association (1991). <https://doi.org/10.2307/1289437>

²⁵ Berne Convention for the Protection of Literary and Artistic Works, September 9, 1886, revised at Paris July 24, 1971, 1161 U.N.T.S. 3 [hereinafter BERNE CONVENTION].

²⁶ BERNE CONVENTION, *supra* note 25.

²⁷ World Intellectual Property Office (WIPO). *Conversation on Intellectual Property (IP) and Artificial Intelligence (AI): Revised Issues Paper on Intellectual Property Policy and Artificial Intelligence- Second Session*, WIPO/IP/AI/2/GE/20/, December 13, 2019.

One of the main challenges of copyright law is balancing the rights of creators with the interests of the public.²⁸ On one hand, the overly restrictive copyright laws hinder access to information and stifle creativity, while insufficient protection can undermine the economic incentives for creators.²⁹ Furthermore, the digital age has introduced new challenges for copyright law, particularly with respect to the distribution and reproduction of digital content.³⁰ The ease of copying and sharing digital works has led to widespread infringement and has necessitated new legal frameworks to address these issues.³¹

Copyright Infringement

Since a copyright holder possesses certain exclusive rights to their intellectual property under the law, any actions taken without his permission that transgresses these rights constitute “copyright infringement.”³² These actions include (1) reproduction of the work or substantial portion of the work, (2) dramatization, translation, adaptation, abridgment, arrangement or other transformation of the work, (3) the first public distribution of the original and each copy of the work by sale or other forms of transfer of ownership, (4) rental of the original or a copy of an audiovisual or cinematographic work, (5) public display of the original or a copy of the work, (6) public performance of the work, (7) and other communication to the public of the work, among others.³³

²⁸ Christophe Geiger. *Copyright as an Access Right, Securing Cultural Participation Through the Protection of Creators' Interests*. (2015): 73-109. <https://doi.org/10.22459/WIWCRC.01.2017.03>

²⁹ *Id.* at 97.

³⁰ Milagros del Corral. *Copyright in the Digital Environment*. *International Information & Library Review*, 29 (1997): 201-204. <https://doi.org/10.1006/IILR.1997.0048>

³¹ *Id.*

³² Calvin N. Mooers. *Computer Software and Copyright*. *ACM Comput. Surv.*, 7 (1975): 45-72. <https://doi.org/10.1145/356643.356647>

³³ An Act Prescribing the Intellectual Property Code and Establishing the Intellectual Property Office, providing for its Powers and Functions and for other purposes, Republic Act no. 8293, as amended, § 177. (June 6, 1997) [hereinafter INTELLECTUAL PROPERTY CODE OF THE PHILIPPINES].

In essence, copyright infringement is piracy, an act that intrudes upon a domain rightfully owned and occupied by a copyright owner.³⁴ To succeed in copyright infringement claims, a plaintiff must prove that (1) he possess a valid copyright for his work, (2) the defendant engaged in an unauthorized copying of the original work, and (3) copying was substantial to constitute an improper or unlawful appropriation that violates a statutory right.³⁵ Intent is irrelevant in the analysis. Hence, even if the defendant genuinely believes that he is not infringing, he could still be held liable if the elements of infringement were established.³⁶

In each case, a copyright owner may choose to bring a lawsuit to recover damages for the infringement of their copyright privileges.³⁷ Although small acts of infringement are typically not brought to court, there are instances where copyright owners pursue legal actions due to its damages to their commercial interests or where such infringing acts would establish a habit of infringement that may jeopardize their copyright or commercial interests in the long run.³⁸ This has been notably evident in the case of copyright in musical compositions³⁹ especially in the Philippines where publishers exert pressure to halt the unauthorized copying and public performances of their musical compositions.

In the case of *Filipino Society of Composers, Authors and Publishers, Inc. (FILSCAP) v. Anrey, Inc.*, the petitioners succeeded in preventing the unauthorized public performance of the copyrighted songs that are included in its repertoire.⁴⁰ The court ruled in its favor and held that the act of playing radio broadcasts containing copyrighted music

³⁴ Pacita I. Habana, Alicia L. Cinco and Jovita N. Fernando v. Felicidad C. Robles and Goodwill Trading Co., Inc., G.R. No. 131522, July 19, 1999.

³⁵ HABANA, ET. AL V. ROBLES, *supra* note 34.

³⁶ Jessica Gillotte. *Copyright Infringement in AI-Generated Artworks* (June 2020). UC Davis Law Review, Vol. 53, No. 5, 2020, Available at SSRN: <https://ssrn.com/abstract=3657423>

³⁷ MOOERS, *supra* note 32.

³⁸ *Id.* at 52.

³⁹ *Id.*

⁴⁰ *Filipino Society of Composers, Authors and Publishers, Inc. (FILSCAP) v. Anrey, Inc.*, G.R. No. 233918, August 09, 2022.

through the use of loudspeakers (radio-over-loudspeakers) is, in itself, a performance which constitutes infringement of the petitioner's copyright over its songs.⁴¹ In the landmark case, *Pearl and Dean (Philippines), Inc. v. Shoemart, Inc.*, the Supreme Court of the Philippines ruled in favor of the petitioners, establishing the principle that copyright infringement occurs when the economic rights of the copyright owner are violated.⁴² This case set a precedent in the protection of audiovisual works and reinforced the importance of recognizing economic rights in copyrighted material.⁴³

The Supreme Court likewise underscored the strict enforcement of copyright protections in broadcasting and educational contexts in the case of *Filipina Broadcasting Network, Inc. v. Ago Medical and Educational Center-Bicol Christian College of Medicine*.⁴⁴ It ruled that even educational institutions must respect copyright laws, emphasizing the necessity of obtaining permission before using copyrighted works.⁴⁵ In the case of *Reyes v. AGB Nielsen Media Research Philippines*, the Court also highlighted the importance of proving ownership and the specific acts of infringement.⁴⁶ This case is significant for its detailed examination of what constitutes infringement and the evidence required to support such claims.⁴⁷

The body of case law in the Philippines on copyright infringement reveals a strong commitment to protecting the rights of creators while adapting to the challenges posed by new technologies. The decisions in key cases underscore the importance of respecting both economic and moral rights. These cases likewise show that copyright infringement transpires at the very moment the infringing acts take place – when the copy is made, a paraphrase is written, when

⁴¹ FILSCAP V. ANREY, *supra* note 40.

⁴² Pearl and Dean (Philippines), Inc. v. Shoemart, Inc., G.R. No. 148222, August 15, 2003.

⁴³ PEARL AND DEAN (PHIL.), INC. V. SHOEMART, INC., *supra* note 42.

⁴⁴ Filipinas Broadcasting Network, Inc. v. Ago Medical and Educational Center-Bicol Christian College of Medicine, G.R. No. 141994, January 17, 2005.

⁴⁵ FILIPINAS BROADCASTING NETWORK, INC. V. AGO MEDICAL AND EDUCATIONAL CENTER, *supra* note 44.

⁴⁶ Reyes v. AGB Nielsen Media Research Philippines, Inc., G.R. No. 176579, October 17, 2008.

⁴⁷ REYES V. AGB NIELSEN, *supra* note 46.

transformation is written out, when the extracts are put into paper, and so on.⁴⁸ Infringements may also occur at the successive stages of producing additional copies and their preparation for distribution to friends, customers, or to the general public.⁴⁹

Understanding copyright infringement in the use of training data

Data are extremely crucial components of AI systems. Each AI model heavily depends on machine learning techniques which utilize big data sets for training and validation.⁵⁰ Hence, a multitude of available data are more likely to produce a better, more authentic, and refined results.⁵¹

Unfortunately, these comprehensive data sets are often composed of copyrighted images, videos, audios, and texts⁵² that are commonly procured without prior authorization.⁵³ This, in turn, raises concerns on whether training data and use of third-party data in AI systems constitute a liability for copyright infringement.⁵⁴

In this regard, this review of related literature shows that there are no clear guidelines in determining the existence of copyright infringement on the use of training data in AI systems. However, scholars have divided views concerning the liability of AI actors in using the said data sets to produce AI-generated outputs.

On one hand, there are scholars who contended that the use of copyrighted materials without the authorization of the original author constitutes copyright infringement and a violation of his intellectual

⁴⁸ MOOERS, *supra* note 32.

⁴⁹ *Id.* at 52.

⁵⁰ AHUJA, *supra* note 15.

⁵¹ *Id.*

⁵² Mark A. Lemley and B. Casey. *Fair Learning*. Intellectual Property: Other eJournal (2020). <https://doi.org/10.2139/ssrn.3528447>

⁵³ Benjamin Sobel. *A Taxonomy of Training Data: Disentangling the Mismatched Rights, Remedies, and Rationales for Restricting Machine Learning*. Electronic (2020). <https://doi.org/10.2139/ssrn.3677548>

⁵⁴ Rustam Mutallimzada. *Liability for Copyright Infringement: An Investigation of the Legal Use of Trained Artificial Neural Networks in the Context of Copyright Law*. Master Thesis, Lund University, 2020.

property rights. On the other hand, the limitations imposed upon the use of copyrighted works as part of the AI's data set encourages bias generation of AI reports and hampers an accurate production of outputs.

Defining copyright infringement in the use of training data

According to Bonadio and McDonagh (2020), an infringement may occur in both the inputs to and outputs of AI since extensive number of copyrighted works fall within the net of the creative AI machines.⁵⁵ As to inputs, the risk of infringing may arise whenever an AI application needs to reproduce the data during the learning process.⁵⁶ Conversely, final outputs produced are considered as adaptation of the works assessed by the machines.⁵⁷

Infringement also occurs when a substantial part of the copyright material is duplicated.⁵⁸ Mutallimzada (2020) asserted that a substantial similarity of the copied work must be identified as the author's original work for an infringement to exist.⁵⁹ This argument is supported by Bonadio and McDonagh (2020) who concluded that the recognition of the final product is a means to identify copyright infringement.⁶⁰ They argued that copyright infringement is less likely to occur if the original upstream work has been changed and transformed to produce marginal similarities in the AI-generated outputs.⁶¹ However, copyright infringement is likely to materialize if the final product includes identifiable and non-secondary elements of the original work.⁶²

⁵⁵ Enrico Bonadio and Luke McDonagh. *Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity*. ERPN: Intellectual Property (2020).

⁵⁶ BONADIO & MCDONAGH, *supra* note 55.

⁵⁷ *Id.* at 8.

⁵⁸ MUTALLIMZADA, *supra* note 54.

⁵⁹ *Id.* at 10.

⁶⁰ BONADIO & MCDONAGH, *supra* note 55.

⁶¹ *Id.* at 8.

⁶² *Id.*

In the case of *Designer Guild Limited (DGL) v. Russel Williams Textiles Limited (RWT)*, the court ruled that RWT infringed upon the *Ixia* design belonging to DGL when it copied the substantial part of the latter in producing its *Marguerite* design.⁶³ This implies that copyright infringement is not concerned about the overall appearance of one's work, but the derivation of one's original work to produce a colorable imitation of the same.⁶⁴

Aside from the copying of the substantial part of the original work, the nature of the reproduction was also considered to determine the existence of copyright infringement. In the case of *Hawkes & Son (London) Ltd v. Paramount Film Service Ltd.*, the use of a 20-second snippet in the newsreel of a 4-minute full march belonging "Colonel Bogey March" is considered as an infringement since the march is easily distinguishable.⁶⁵ Copyright infringement is likewise present in the case of *HRH Prince of Wales v. Associated Newspapers* when the latter used 15% of the unauthorized extracts from Prince of Wales's journal to be published on The Mail since the Prince's opinions were of highest interest in the newspaper's readers.⁶⁶

The complex and broad-reaching AI technologies like robots are similarly prone to committing copyright infringement, especially if they are designed to store and share data that are likely protected by copyright.⁶⁷ Robots, for instance, are programmed to store pictures, videos, a piece of digital information and other data on its hard drive which can be reproduced, held, and processed.⁶⁸ Taking, storing, and sharing these information may constitute possible copyright infringement.⁶⁹

Xu Xinming (2017, as cited in Mutallimzada, 2020) added that copyright infringement may exist if the content collected is without

⁶³ *Designers Guild v Russell Williams* [2001] 1 All ER 700.

⁶⁴ *Id.*

⁶⁵ *Hawkes v Paramount Film services* [1934] Ch 593.

⁶⁶ *HRH Prince of Wales v Associated Newspapers* [2008] IP & T 583.

⁶⁷ Aviv Hertzal Gaon. *Artificially Intelligent Copyright: Rethinking Copyright Boundaries*. Dissertation, York University, 2019.

⁶⁸ GAON, *supra* note 67.

⁶⁹ *Id.* at 161.

clear authorship.⁷⁰ To illustrate, “Clearview AI” is a new facial recognition software engine with the capability to extract and match more than three billion images from various websites and social media platforms.⁷¹ The use of said software raised significant privacy issues and debatable questions concerning copyright infringement of AI-generated work since scraping and retaining images in Clearview AI’s database can potentially result to infringement of copyright-protected images.⁷²

Third-party use of data may likewise constitute infringement with the use of artificial intelligence that can be instructed to conduct certain tasks and produce defined outputs. In order to build an AI application that generates outputs based on trained data, it is necessary to have access and to secure authorization on the use of data for training purposes.⁷³ For instance, to produce an academic essay on Philippine history, access to numerous books, journals, archives, and relevant articles is needed. Considering that these data are protected under copyright law, any unauthorized use of such data without the intellectual property right holder’s permission constitutes copyright infringement.⁷⁴

Reproducing copyrighted work is likewise deemed as the most infringing act which commonly occurs in the reproduction of literary, artistic, and musical work for commercial purposes. In this case, the ones authorizing the infringing act by way of facilitating performance within the infringing work might also be included in the proceedings as a joint tortfeasor.⁷⁵

⁷⁰ MUTALLIMZADA, *supra* note 54.

⁷¹ *Id.* at 4.

⁷² Teresa Scassa, ‘How Clearview AI Could Violate Copyright Law: Clearview AI scraped countless copyright-protected images from social media sites to develop a commercial facial recognition technology. Could copyright law be used to dismantle its service?’, (CIGI, 10 March 2020), <https://www.cigionline.org/articles/how-clearview-ai-could-violate-copyright-law> (last accessed 10 December 2023).

⁷³ MUTALLIMZADA, *supra* note 54.

⁷⁴ *Id.* at 7.

⁷⁵ *Id.* at 15.

CHAPTER II

AI LEGISLATION AROUND THE GLOBE: LIABILITIES OF AI ACTORS IN UNAUTHORIZED USES OF COPYRIGHTED MATERIALS

The review of copyright-related provisions in AI legislation across countries shows that each nation has its unique approach to addressing copyright concerns in the context of AI. This analysis includes laws and legal frameworks from the United States, European Union, United Kingdom, Japan, China, Singapore, Australia, and Philippines.

United States

Several federal laws regarding AI have been enacted over the past few Congresses in the United States, either as individual bills or as specific AI-related provisions within the larger acts. Among these laws are the expansive National Artificial Intelligence Initiative Act of 2020 (NAII)⁷⁶ and the National Defense Authorization Act (NDAA).⁷⁷

On one hand, NAII establishes the American AI Initiative and provides guidelines for AI research, development, and assessment activities at federal science agencies.⁷⁸ Individual agencies have also been statutorily directed to engage in activities that promote the use of AI across the federal government.⁷⁹ On the other hand, NDAA has

⁷⁶ A bill to establish the National Artificial Intelligence Initiative, and for other purposes. House of Representatives 6216, (Mar. 12, 2020) (US), <https://www.congress.gov/116/bills/hr6216/BILLS-116hr6216ih.pdf>

⁷⁷ An act to authorize appropriations for fiscal year 2024 for military activities of the Department of Defense and for military construction, and for defense activities of the Department of Energy, to prescribe military personnel strengths for such fiscal year, and for other purposes. H.R. 2670 (Dec. 22, 2023) (US), <https://www.congress.gov/118/plaws/publ31/PLAW-118publ31.pdf> [hereinafter NATIONAL DEFENSE AUTHORIZATION ACT].

⁷⁸ HOUSE OF REPRESENTATIVES 6216, *supra* note 76 at § 101(a).

⁷⁹ *Id.* at § 103(c)

already included provisions on AI in defense, national security, and intelligence.⁸⁰

NAII contains provisions on machine learning particularly on use of funds for “managing and making available to researchers accessible, curated, standardized, secure, and privacy protected data sets from public and private sectors for the purposes of training and testing artificial intelligence systems and for research using artificial intelligence systems.”⁸¹

In order to foster best practices and voluntary standards for trustworthy artificial intelligence systems, NAII likewise provides for “data management and techniques to increase the usability of data, including strategies to systematically clean, label, and standardize data into forms useful for training artificial intelligence systems and the use of common, open licenses.”⁸² The same Act also directs the Director to collaborate with the public and private sectors in developing “guidance to facilitate the creation of voluntary data sharing arrangements between industry, federally funded research centers, and Federal agencies for the purpose of advancing artificial intelligence research and technologies.”⁸³

Europe

The European Parliament has released the final draft of the Artificial Intelligence Act (AI Act) in April 2024 which lays the foundations for the regulation of AI in the European Union.⁸⁴ The AI Act aims to establish a regulatory oversight over the development and

⁸⁰ NATIONAL DEFENSE AUTHORIZATION ACT, *supra* note 77 at § 1541.

⁸¹ *Id.* at § 3(A).

⁸² *Id.* at III § 302(a)(1)(C).

⁸³ § 301(c)

⁸⁴ European Commission, Artificial Intelligence Act: European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (2019-2024) (EU), https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

utilization of various applications based on its classified risks.⁸⁵ These risks include (1) unacceptable risk which is prohibited,⁸⁶ (2) high-risk AI systems that are regulated,⁸⁷ (3) limited risk AI systems which are subject to lighter transparency obligations,⁸⁸ and (4) minimal risk that are unregulated.⁸⁹

The AI Act contains provisions on General Purpose AI (GPAI) which requires all GPAI model to prepare technical documentation, provide instructions for use, adhere to the copyright directive, and publish a summary of training content.⁹⁰ Under the AI Act, providers of GPAI models under a free and open license likewise need to comply with copyright requirements and publish the training data summary, unless they pose a systemic risk.⁹¹ Additionally, the AI Act directs all providers of GPAI models that present systemic risk to conduct model evaluations, perform adversarial testing, track and report serious incidents, and ensure cybersecurity protections.⁹²

United Kingdom

In February 2024, as part of its effort to continuously improve the National AI Strategy,⁹³ the UK Government has unveiled its response to the White Paper consultation on AI regulation which reaffirms its “pro-innovation” stance that aim to balance innovation with safety by applying its existing technology-neutral regulatory framework to AI.⁹⁴

⁸⁵ ARTIFICIAL INTELLIGENCE ACT, *supra* note 84.

⁸⁶ ARTIFICIAL INTELLIGENCE ACT, *supra* note 84 at § 46.

⁸⁷ *Id.* at § 52.

⁸⁸ *Id.* at § 53.

⁸⁹ *Id.*

⁹⁰ *supra* note 84 at § 52(2).

⁹¹ *Id.* at § 104.

⁹² *Id.* at § 3(55)(1)(a).

⁹³ Secretary of State for Digital, Culture, Media and Sport by Command of Her Majesty. (2021). *National AI Strategy*. In HM Government (ISBN 978-1-5286-2894-5). Office for Artificial Intelligence.

https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf (last accessed June 2, 2024)

⁹⁴ Department for Science, Innovation, and Technology [DSIT]. *A pro-innovation approach to AI regulation: government response: Presented to Parliament by the Secretary of State for*

This contains provisions in enabling the AI-based R&D and innovation in the creative industries,⁹⁵ and prioritizing trust and transparency among its stakeholders.⁹⁶

In terms of intellectual property, UK is geared towards protection of copyrighted contents by the creative industries through advocating transparency from AI developers in relation to data inputs and proper attribution of outputs.⁹⁷

Japan

In April 2024, Japan has also introduced the Draft AI Guidelines for Businesses⁹⁸ and the Hiroshima AI Process Comprehensive Policy Framework⁹⁹ as part of its efforts to issue guidelines governing artificial intelligence. These guidelines contain provisions on appropriate data input measures and protection to safeguard one's personal data and intellectual property.¹⁰⁰ This particularly implements machine learning techniques for transparency and privacy protection like testing and finetuning leaks of sensitive data.¹⁰¹ It likewise introduces appropriate safeguards to respect privacy and

Science, Innovation and Technology by Command of His Majesty on 6 February 2024.
<https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response#summary-of-consultation-evidence-and-government-response>.
February 6, 2024. (last accessed May 29, 2024).

⁹⁵ DSIT, *supra* note 94 at § 27.

⁹⁶ *Id.* at § 31.

⁹⁷ *Id.* at § 28.

⁹⁸ Ministry of Internal Affairs and Communications Ministry of Economy, Trade and Industry. AI Guidelines for Business Ver1.0 (April 19, 2024) (JP).
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240419_9.pdf

⁹⁹ Hiroshima AI Process Comprehensive Policy Framework, May 2, 2024.
<https://www.soumu.go.jp/hiroshimaaiprocess/en/index.html#:~:text=After%20continuous%20discussions%20including%20at,code%20of%20conduct%20aimed%20at>.

¹⁰⁰ HIROSHIMA PROCESS, *supra* note 99 at § XI.

¹⁰¹ *Id.* at § 2(xx).

intellectual property rights, including those related to copyright protected content.¹⁰²

Developments on AI are likewise integrated on the copyright law of Japan. Its Copyright Act¹⁰³ promotes data utilization in machine learning.¹⁰⁴ Its amendment clarified that downloading or processing the data using the internet or any other methods for creating AI models does not constitute copyright infringement.¹⁰⁵

China

In its attempt to establish the world's earliest and most comprehensive AI regulations,¹⁰⁶ China has released a draft of its Artificial Intelligence Law of the People's Republic of China through the Center for Security and Emerging Technology (CSET) in May 2024.¹⁰⁷ Article 24 of the said law provides for the reasonable use of data.¹⁰⁸ It specifically states that "when an AI developer uses the copyrighted data of others for model training, if the use is different from the original purpose or function of the data and does not affect the normal use of the data or unreasonably harm the legitimate rights

¹⁰² *Id.*

¹⁰³ Copyright Act of Japan, Act No. 48 of 1970, as amended by Act No. 30 of 2018. [hereinafter COPYRIGHT ACT OF JAPAN].

¹⁰⁴ COPYRIGHT ACT OF JAPAN, *supra* note 103 at § 30-4.

¹⁰⁵ *Id.* at § 30-4(ii)(iii).

¹⁰⁶ Matt Sheehan and Carnegie Endowment for International Peace [Carnegie]. *China's AI Regulations and How They Get Made*. Carnegie Endowment for International Peace Publications Department, July 10, 2023. https://carnegie-production-assets.s3.amazonaws.com/static/files/202307-Sheehan_Chinese%20AI%20gov-1.pdf. (last accessed May 29, 2024).

¹⁰⁷ CSET. *Artificial Intelligence Law of the People's Republic of China (Draft for Suggestions From Scholars)* Center for Security and Emerging Technology. Center for Security and Emerging Technology, May 2, 2024. <https://cset.georgetown.edu/publication/china-ai-law-draft/>.

¹⁰⁸ ARTIFICIAL INTELLIGENCE LAW OF THE PEOPLE'S REPUBLIC OF CHINA, *supra* note 107 at § 24.

and interests of the data's owner, such use is a reasonable use of data."¹⁰⁹

Singapore

In January 2024, Singapore's Infocomm Media Development Authority (IMDA), in collaboration with the AI Verify Foundation launched a public consultation for its draft of the Model AI Governance Framework for Generative AI.¹¹⁰ While the draft aids various organizations in grasping key policy implications, it served more as a discussion paper rather than prescribing specific practices for organizations to follow when implementing generative AI solutions.

The model recognizes the importance of data as a core element in model and application development. Since training requires large volume of data, it typically includes publicly available data and copyrighted materials which becomes part of the web-scraped datasets.¹¹¹ In order to address such cases, the model emphasized the need to balance competing concerns and ensure the fair treatment in pragmatic manner.¹¹²

The model went as far as highlighting the concerns of both the copyright owners and companies involved in AI development. On one hand, entities engaged in AI development argued that "text and data mining, and machine learning do not infringe copyright because training does not involve the copying and use of the creative

¹⁰⁹ *Id.*

¹¹⁰ Marcus Evans, Wilson Ang, Jeremy Lua, and Jeremiah Chew. *Singapore proposes Governance Framework for Generative AI*. Data Protection Report. February 4, 2024. <https://www.dataprotectionreport.com/2024/02/singapore-proposes-governance-framework-for-generative-ai/>.

¹¹¹ AI Verify Foundation and Infocomm Media Development Authority [IMDA]. *Propose Model AI Governance Framework for Generative AI: Fostering a Trusted Ecosystem*, January 16, 2024. https://aiverifyfoundation.sg/downloads/Proposed_MGF_Gen_AI_2024.pdf. (last accessed May 30, 2024).

¹¹² IMDA, *supra* note 111 at 8.

expression in works.”¹¹³ On the other hand, copyright owners have “requested remuneration for use of their works to train models...as such systems may compete with them and impact their livelihood.”¹¹⁴ In particular, the copyright owners are advocating for licensing-based solutions “to facilitate text and data mining activities for machine learning, as well as an opt-out system for copyright owners from statutory exceptions for text and data mining and machine learning activities to avoid duly impinging on their commercial interests.”¹¹⁵

Australia

In January 2024, the Australian Government has published its interim response to public consultation regarding AI regulation.¹¹⁶ This report outlines the measures that the government will take to ensure that AI is designed, developed, and deployed safely and responsibly.¹¹⁷ The report addressed the risks posed by the inaccuracies in system design and biases in training data.¹¹⁸ To combat inaccurate and unfair outcomes resulting from errors in data inputs, the Australian government recommends transparency and accountability in the use of datasets and training data.

The report specifies the need for public reporting of data where the model is trained on and sharing information on data processing and testing.¹¹⁹ Accountability could include having designated roles

¹¹³ *Id.*

¹¹⁴ *Id.*

¹¹⁵ *Id.*

¹¹⁶ Australian Government: Department of Industry, Science and Resources. *Supporting responsible AI: discussion paper*, January 17, 2024. <https://consult.industry.gov.au/supporting-responsible-ai>. (last accessed May 30, 2024).

¹¹⁷ AUSTRALIAN GOVERNMENT, *supra* note 116.

¹¹⁸ Commonwealth of Australia. *Safe and Responsible AI in Australia Consultation: Australian Government's interim response*. Australian Government: Department of Industry, Science, and Resources. Commonwealth of Australia, 2024. https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/safe-and-responsible-ai-in-australia-governments-interim-response.pdf. (last accessed May 30, 2024).

¹¹⁹ COMMONWEALTH OF AUSTRALIA, *supra* note 118 at 11.

with responsibility for AI safety and requiring training for developers and deployers of AI products in certain settings.¹²⁰ The same also recommends considering watermarking or similar data provenance mechanisms.¹²¹ Australia likewise initiated steps in integrating AI provisions into its copyright laws, particularly concerning the use of copyrighted contents for training generative AI models and potential remedies for infringement.¹²²

Philippines

The Philippines does not have a specific law which exclusively governs artificial intelligence, but there are existing laws and policies containing provisions that address several AI concerns. The Data Privacy Act of 2012¹²³ sets the standards for data protection and privacy especially in processing large amounts of personal data that are crucial for AI models. The Electronic Commerce Act of 2000¹²⁴ likewise facilitates the legal framework for digital transactions and electronic data while the Cybercrime Prevention Act of 2012¹²⁵ provides for the prevention, investigation, and prosecution of cybercrime offenses.

In terms of copyright, Republic Act 8293 or the Intellectual Property Code of the Philippines¹²⁶ addresses issues related to

¹²⁰ *Id.* at 20.

¹²¹ *Id.* at 21.

¹²² *Id.* at 15.

¹²³ An Act Protecting Individual Personal Information in Information and Communications Systems in the Government and the Private Sector, creating for this purpose a National Privacy Commission, and for other Purposes, Republic Act 10173 (August 15, 2012) [hereinafter DATA PRIVACY ACT OF 2012].

¹²⁴ An Act Providing for The Recognition and Use of Electronic Commercial and Non-Commercial Transactions, Penalties for Unlawful Use Thereof, and other purposes, Republic Act No. 8792, (June 14, 2000) [hereinafter ELECTRONIC COMMERCE ACT].

¹²⁵ An Act Defining Cybercrime, Providing for the Prevention, Investigation, Suppression and the Imposition of Penalties Therefor and for other purposes, Republic Act No. 10175, (September 12, 2012) [hereinafter CYBERCRIME PREVENTION ACT OF 2012].

¹²⁶ An Act Prescribing the Intellectual Property Code and Establishing the Intellectual Property Office, providing for its Powers and Functions and for other purposes,

unauthorized uses of copyrighted materials and protection to an individual's intellectual property. Section 177 of the RA 8293 provides that the copyright or economic rights of the rights holder consist of exclusive right to carry out, authorize, or prevent certain acts.¹²⁷ Some of these acts include "reproduction of the work or substantial portion of the work"¹²⁸ and "rental of the original or a copy of ... a computer program, a compilation of data and other materials...irrespective of the ownership of the original or the copy of which is the subject of the rental."¹²⁹

Table 1 summarizes the relevant provisions and clauses concerning copyright and use of copyrighted materials as training data within various global AI legislations. It specifically covers laws and frameworks from the United States, Europe, the United Kingdom, Japan, China, Singapore, Australia, and the Philippines.

Republic Act no. 8293, as amended, (June 6, 1997) [hereinafter INTELLECTUAL PROPERTY CODE OF THE PHILIPPINES].

¹²⁷ INTELLECTUAL PROPERTY CODE OF THE PHILIPPINES, *supra* note 126 at § 177.

¹²⁸ *Id.* at § 177.1.

¹²⁹ § 177.4.

Table 1. Summary of relevant provisions and clauses concerning copyright and the use of copyrighted materials as training data in various global legislations

COUNTRIES	RELEVANT PROVISIONS / CLAUSES	WHAT IT SAYS ABOUT COPYRIGHT?
<p>United States National Artificial Intelligence Initiative Act of 2020 (NAII)</p>	<ol style="list-style-type: none"> 1. Sec. 3(9) defines machine learning as an application of AI that has the ability to learn and improve on the basis of data.¹³⁰ 2. Sec. 103 (d)(2)(B)(iv) details the responsibility of the Interagency Committee to offer and make accessible datasets for AI research and development.¹³¹ 3. Sec. 201 (b)(3)(A) specifies the financial assistance allocated for organizing datasets that will be used for training and testing AI systems.¹³² 4. Sec. 301 (a)(1)(A) promotes best practices and voluntary standards for trustworthy AI systems. It ensures privacy and security for datasets used in training or testing AI systems.¹³³ 5. Sec. 301 (a)(1)(C) promotes improving data management and techniques to enhance data usability.¹³⁴ 	<p>NAII contains general provisions on facilitating research and development with the use of artificial intelligence. It also includes specific provisions on machine learning and the use of data as inputs for datasets in producing output. However, it does not contain provisions specifically directed to copyright or infringing copyrights of owner or copyright holder.</p> <p>For copyright concerns, it made reference to the Copyright Act and other laws governing copyright.</p>

¹³⁰ HOUSE OF REPRESENTATIVES 6212, *supra* note 76 at § 3(9).

¹³¹ *Id.* at § 103 (d)(2)(B)(iv).

¹³² § 201 (b)(3)(A).

¹³³ § 301 (a)(1)(A)

¹³⁴ § 301 (a)(1)(C)

	<p>6. Sec. 301 (c) directs the Director to develop guidance to promote voluntary data sharing between industries.¹³⁵</p> <p>7. Sec. 301 (c)(2) establishes best practices for datasets that involve human characteristics.¹³⁶</p>	
<p>Europe Artificial Intelligence Act (AI Act)</p>	<p><i>As provided for in its whereas clause:</i></p> <p>1. Paragraph 104 provides that general purpose AI (GPAI) must produce a summary of the training data and have a policy that complies with the Union Copyright Law. GPAI that were released under a free and open-source license and have publicly available parameters, architecture, and usage information may be exempted from transparency requirements.¹³⁷</p> <p>2. Paragraph 105 recognizes the applicability of Directive (EU) 2019/790¹³⁸ in developing and training GPAI models. Directive (EU) 2019/790 allows text and data mining under certain conditions but gives the rights holders an option to reserve their rights to prevent it except for scientific and research purposes. If rights are reserved, AI model providers must obtain authorization for text and data mining.¹³⁹</p>	<p>The AI Act does not specifically provide for the liabilities of the AI actors concerning the unauthorized use of copyrighted materials. However, it set forth the guidelines on the permissible and impermissible actions of AI actors when utilizing both copyrighted and publicly available materials as training data for general purpose AI models.</p> <p>The AI Act also made explicit reference to Directive (EU) 2019/790 for copyright concerns.</p>

¹³⁵

§ 301 (c)

¹³⁶

§ 301 (c)(2)

¹³⁷

ARTIFICIAL INTELLIGENCE ACT, *supra* note 84 at § 104.

¹³⁸

Directive (EU) 2019/790 of the European Parliament and of the Council on Copyright and Related Rights in the Digital Single Market and Amending Directives 96/9/EC and 2001/29/EC of 17 April 2019 (July 5, 2019) [hereinafter DIRECTIVE (EU) 2019/790].

¹³⁹

ARTIFICIAL INTELLIGENCE ACT, *supra* note 84 at § 105.

	<p>3. Paragraph 106 specifies that providers of GPAI must comply with the Union Copyright Laws in establishing policies to identify and respect the reservation of the rights holders to ensure fair competition.¹⁴⁰</p> <p>4. Paragraph 107 facilitates transparency by requiring GPAI models to publish detailed summary of their training data. The summary should be comprehensive but not overly technical so as to help the copyright holders to understand and enforce their interests under the Union Law.¹⁴¹</p> <p>5. Paragraph 108 reiterates the obligation of the developers of the GPAI model to comply with the Union Copyright Law in publishing a summary of their training data, although the AI Office will not verify each piece of training data for copyright compliance.¹⁴²</p> <p>6. Paragraph 109 exempts developers and users of GPAI models for non-professional or scientific research, but they are encouraged to follow the requirements voluntarily. It states that compliance must with the obligations must be proportional to the provider type.¹⁴³</p>	
<p>United Kingdom A pro-innovation approach to AI regulation:</p>	<p>As provided in its report on <i>enabling AI innovation and protecting intellectual property</i>:</p> <p>1. Paragraph 28 addresses the concerns of creative industries and media organizations in the widespread use of their</p>	<p>The National AI Strategy and the report on UK's pro-innovation approach to AI regulation do not contain provisions on the liabilities of AI actors in the unauthorized use of copyrighted materials.</p>

¹⁴⁰ *Id.* at § 106.

¹⁴¹ § 107.

¹⁴² § 108.

¹⁴³ § 109.

<p>government response (A report to further improve the National AI Strategy)</p>	<p>copyrighted materials for training data and the preservation of their ability to control their work. It also put emphasis on the need for access of AI developers to diverse and high-quality datasets in UK.¹⁴⁴</p> <p>2. Paragraph 29 details the effort of the UK government in bringing together copyright holders and AI developers to exchange opinions, but the group was not able to reach a consensus on an effective voluntary code.¹⁴⁵</p> <p>3. Paragraph 30 engages an effective approach to support the growth of the creative sectors while ensuring that AI development complements human creativity, innovation, and the delivery of reliable information.¹⁴⁶</p> <p>4. Paragraph 31 emphasizes the reliance of the UK government on trust and transparency among all the parties involved, especially on how AI developers use the copyrighted data and how they attribute copyright holders to their outputs.¹⁴⁷</p>	<p>However, they outline future plans to address copyright-related concerns, such as extensive use of copyrighted materials for training AI models, enhancing creativity and innovation, ensuring the delivery of reliable information, and promoting transparency in data usage.</p> <p>UK’s AI laws did not make any explicit reference to its copyright law, but it emphasized that “any new rules for AI should not contradict or duplicate existing laws”¹⁴⁸ and that they will continue to examine questions related to accountability and liability, specifically investigating the extent to which the existing laws and regulators can reach through the developers of advanced general-purpose systems.¹⁴⁹</p>
---	---	--

¹⁴⁴ DSIT, *supra* note 94 at 16 § 28.

¹⁴⁵ *Id.* at § 29.

¹⁴⁶ § 30.

¹⁴⁷ § 31.

¹⁴⁸ DSIT, *supra* note 94 at 39 § 113.

¹⁴⁹ *Id.* at 52 § 180.

<p>Asia (<i>Japan, China, Singapore, Australia, Philippines</i>)</p>		
<p>Japan Copyright Act of Japan (as amended)</p>	<ol style="list-style-type: none"> 1. Article 2 (x) defines database as a collection of data that is organized in a systematic manner to allow computer-based searching.¹⁵⁰ 2. Article 12(1) specifies that a compilation is protected as a creative work if its selection.¹⁵¹ 3. Article 12-2(1) states that database is protected as a work if its selection or systematic organization of information demonstrates creative effort.¹⁵² 4. Article 12-2(2) provides that the protection outlined in the previous paragraph does not impact the rights of the author of a work included in the referenced database.¹⁵³ 5. Article 30-4 states that it is permissible to exploit a work in any situation where the intent is not for personal enjoyment or to let others enjoy the work’s ideas or sentiment of another. However, such permission would not apply if the use of would unreasonably harm the interests of the copyright owner.¹⁵⁴ 6. Article 30-4(ii) specifies one of the instances permitted under Article 30-4. This includes the use of data for data 	<p>The AI laws of Japan do not specifically address copyright issues, but it made reference to its Copyright Act to address copyright concerns relating to AI technologies.</p> <p>Japan’s Copyright Act permits the use of copyrighted materials as training data, provided that it does not infringe on the rights of the copyright holders. It grants broad permission for utilizing the work of other under the following conditions: (1) its use is not for personal enjoyment of the work’s ideas or sentiments, (2) it facilitates the incidental use of work, (3) it is employed for data processing and analysis, (4) its use does not significantly harm the interests of the copyright holders, and (5) the materials used are not made available illegally.</p>

¹⁵⁰ COPYRIGHT ACT OF JAPAN, *supra* note 103 at § 2(x).

¹⁵¹ *Id.* at § 12(1).

¹⁵² § 12-2(1).

¹⁵³ § 12-2(2).

¹⁵⁴ § 30-4.

	<p>analysis (extraction, comparison, classification, other statistical evaluation of language, sound, images, etc.).¹⁵⁵</p> <p>7. Article 30-4(iii) mentions that exploitation of one’s work is likewise permissible if it is used in computer data processing or in a manner that does not involve the work being perceived by human senses.¹⁵⁶</p> <p>8. Article 47-4(1) states that a person may use a work available on a computer in instances where the purpose is to facilitate the incidental use of the work on a computer for smooth or efficient operation. However, this does not apply if such use would unreasonably cause harm to the interests of the copyright owners.¹⁵⁷</p> <p>9. Article 47-5(1) provides that a person engaged in activities that create new knowledge or information through computerized data processing may use publicly available or presented work as needed for those activities. The use must only be minor (involves small part of the work only) and does not significantly harm the interests of the copyright owners. This does not apply to work that is made available illegally.¹⁵⁸</p>	
<p>China Artificial Intelligence Law of the People’s Republic of China</p>	<p>1. Article 10 Development of Infrastructure – provides for the enhancement of the AI infrastructure by effectively managing and allocating national computing power resources and supporting the creation of open public data</p>	<p>China’s AI law specifically addresses copyright issues, allowing the use of copyrighted materials as training data. This is permitted as long as the resulting work serves different purpose from the</p>

155 § 30-4 (ii).
156 § 30-4 (iii).
157 § 47-4(1).
158 § 47-5(1).

	<p>platforms that are essential for the growth of the AI industry.¹⁵⁹</p> <p>2. Article 23 Protection of IPR – mentions that the State establishes and improves rules protecting intellectual property rights (IPR) related to training data, algorithms, and AI-generated content, as well as IPR resulting from the use of AI in creation, invention, and industrial design.¹⁶⁰</p> <p>3. Article 24 Reasonable Use of Data – specifies that AI developers’ use of copyrighted data for model training is considered reasonable when it is used in a way that differs from its original purpose, it does not interfere with its normal use, and does not unreasonably harm the data owner’s rights. In such case the AI developer can use the data without paying the owner or needing permission, but they must clearly mark the data source.¹⁶¹</p> <p>4. Article 36 IPR of AI-Generated Content – articulates that when AI-generated content meets the conditions of the copyright law, it can be protected based on the user’s contribution, but the user must disclose that the content is primarily generated by AI.¹⁶²</p> <p>5. Article 45 Data Quality – states that AI developers must implement effective measures to improve the quality of data by ensuring its veracity, accuracy, objectivity, and</p>	<p>original, does not interfere with the normal use of the data, and does not harm the legitimate rights and interests of the copyright holders.</p>
--	---	--

¹⁵⁹ ARTIFICIAL INTELLIGENCE LAW OF THE PEOPLE’S REPUBLIC OF CHINA, *supra* note 107 at § 10.

¹⁶⁰ *Id.* at § 23.

¹⁶¹ § 24

¹⁶² § 36

	diversity. Appropriate action is also required of them when they discovered any information in the training data that is prohibited by laws and regulations. ¹⁶³	
Singapore Model AI Governance Framework for Generative AI	<ol style="list-style-type: none"> 1. Singapore recognizes the need to balance user protection with the promotion of innovation. It argues that AI concerns should be addressed practically and holistically.¹⁶⁴ 2. The framework also specifies that when AI models use potentially contentious data like personal data and copyrighted materials, the businesses should be given clarity, ensure fair treatment, and handle the issue pragmatically.¹⁶⁵ 3. Ex Post Safety Nets - The framework highlights the developers' implicit recognition of its accountability for the training data used and the applications of their models through third-party copyright claims related to their AI products and services.¹⁶⁶ 4. Balancing copyright with data accessibility - Singapore recommends that solutions to copyright concerns relating AI should be balanced and align with market realities.¹⁶⁷ 	<p>The Singaporean AI framework does not contain explicit regulations concerning the liabilities of AI actors in an unauthorized use copyrighted materials, but it has already initiated several steps in promoting balanced user protection and innovation.</p> <p>While there is no direct reference made to the copyright law, Singapore advocates for balanced copyright protection and data accessibility.</p>

¹⁶³ § 45

¹⁶⁴ IMDA, *supra* note 111 at 3.

¹⁶⁵ *Id.* at 4.

¹⁶⁶ *Id.* at 22.

¹⁶⁷ *Id.* at 9.

<p>Australia Safe and Responsible in Australia</p>	<ol style="list-style-type: none"> 1. In its report, Australia mentioned that businesses and individuals who develops and uses AI are already subjected to various Australian laws, although there is still a need to update its copyright law in terms of how creative content can be utilized in training generative AI models, along with potential solutions for addressing copyright infringement.¹⁶⁸ 2. In its effort to strengthen its laws, Australia also reported that there are ongoing research and consultations on the implications of AI for copyright and broader IP law through the AI Working Group of the IP Policy Group.¹⁶⁹ 3. In order to address copyright AI concerns, the Australian government advocated for transparency and accountability on the use of data. Transparency measures include (a) informing users when AI systems are in use or when content is AI-generated, potentially through labelling or watermarking, (b) public reports on limitations, capabilities, and appropriate uses of AI systems, and (c) disclosures about the data used for training models and details on data processing and testing.¹⁷⁰ 4. Accountability measures, on the other hand, include assigning specific roles dedicated to AI safety and mandating training for developers and deployers of AI products in certain settings.¹⁷¹ 	<p>The Australian framework lacks specific provisions regarding the liability of AI actors for unauthorized use of copyrighted materials. Nonetheless, it acknowledges the necessity of amending existing copyright laws to address issues related to using copyrighted materials for training generative AI models and finding solutions for copyright infringement.</p> <p>Additionally, the framework advocates for transparency and accountability to ensure the responsible use of data in the training, processing, and development of AI programs.</p>
---	---	---

¹⁶⁸ COMMONWEALTH OF AUSTRALIA, *supra* note 118 at 15.

¹⁶⁹ *Id.* at 22.

¹⁷⁰ *Id.* at 20.

¹⁷¹ *Id.*

	<p>5. Australia likewise advocated for the use of voluntary watermarking or other similar data provenance methods for AI use.¹⁷²</p>	
<p>Philippines Intellectual Property Code of the Philippines (RA 8293)</p>	<ol style="list-style-type: none"> 1. Section 177 defines the copyright or economic rights to include exclusive rights to carry out, authorize, or prevent certain acts.¹⁷³ 2. Section 177.1 specifies one of the acts mentioned in the previous paragraph to include “reproduction of the work or substantial portion of the work.”¹⁷⁴ 3. Section 177.4 likewise mentions the rights holder’s copyright over “rental of the original or a copy of... a computer program, a compilation of data and other materials... irrespective of the ownership of the original or the copy which is the subject of the rental.”¹⁷⁵ 4. Section 187 provides that “the private reproduction of a published work in a single copy, where the reproduction is made by a natural person exclusively for research and private study, shall be permitted, without the authorization of the owner of copyright in the work.”¹⁷⁶ 5. Section 187.2 enumerated the exceptions to Section 187. This includes the reproduction of the architectural works, 	<p>The Philippines does not have an existing law governing artificial intelligence. Its copyright concerns are addressed by the Intellectual Property Code of the Philippines (RA 8293).</p> <p>RA 8293 defines copyright as the economic right accorded to the rights holders over their works, including computer programs and compilation of data. Although the Code assures the protection of both economic and moral rights of the rights holders, it does not address the unauthorized use of copyrighted materials by AI applications or any associated liabilities for its use as training data.</p>

¹⁷² *Id.* at 21.

¹⁷³ INTELLECTUAL PROPERTY CODE OF THE PHILIPPINES, *supra* note 126 at § 177.

¹⁷⁴ *Id.* at § 177.1.

¹⁷⁵ § 177.4.

¹⁷⁶ § 187

	<p>entire books or substantial of the same, musical works in graphic form by reprographic means, compilations of data and other materials, computer programs, or any work where reproduction would unreasonably conflict with normal exploitation or prejudice the legitimate interests of the author.¹⁷⁷</p> <p>6. Section 173 considers the compilation of data as derivative works. "Other materials which are original by reason of the selection or coordination or arrangement of their contents" are likewise considered as derivative works.¹⁷⁸</p> <p>7. Section 175 provides that no protection shall extend to "mere data... even if they are expressed, explained, illustrated, or embodied in a work."¹⁷⁹</p>	
--	---	--

¹⁷⁷ § 187.2

¹⁷⁸ § 173.

¹⁷⁹ § 175.

Table 1 shows that each country has its unique approach to addressing copyright concerns within the context of AI legislation. United States focuses on facilitating research and development in artificial intelligence while United Kingdom advocates for pro-innovation approaches to AI regulation. Europe provides guidelines for AI actors regarding the permissible and impermissible uses of copyrighted materials as training data while China and Japan permit the use of copyrighted materials for training AI under certain conditions. Australia and Singapore acknowledged the need to amend their existing copyright laws to address issues related to copyrighted materials for AI training, emphasizing transparency and accountability, and balanced copyright protection and data accessibility. On the other hand, due to lack of AI legislation or framework, the Philippines continue to depend on the limited provisions provided under its Copyright law.

Overall, while some other countries have specific provisions addressing copyright concerns within their AI legislation, others rely on their existing copyright laws and frameworks to handle copyright issues related to AI technologies. On one hand, Singapore and Australia do not have any AI legislation, instead both rely on their existing frameworks to manage copyright issues. On the other hand, United States, European Union, and Japan have already drafted their respective AI legislations, but they still reference their domestic copyright laws in addressing copyright concerns related to AI technologies. In contrast, China is the sole country which explicitly incorporated provisions regarding the unauthorized uses of copyrighted materials as part of the training data of AI models. This finding highlights a gap that needs to be addressed in crafting future regulatory developments.

The liabilities of the AI actors are likewise not extensively addressed in any AI law or framework covered because majority of the AI legislation allow or promote the utilization of copyrighted materials as part of the training data, either explicitly or implicitly, under specific conditions.

Various permissible acts concerning the use of copyrighted materials in AI development can likewise be inferred from the provisions of the cited AI legislations and frameworks that balance innovation with intellectual property protection.

This includes (1) exemption from transparency requirements of general-purpose AI models that were released under free and open-source license,¹⁸⁰ (2) text and data mining,¹⁸¹ (3) use of copyrighted data for non-professional and scientific research,¹⁸² (4) exploitation of work for non-personal enjoyment,¹⁸³ (5) exploitation of work used in computer data processing,¹⁸⁴ (6) computer data used to facilitate incidental use for smooth and efficient operation,¹⁸⁵ (7) use for training data to create new knowledge through computerized data processing,¹⁸⁶ and (8) the use of copyrighted data for model training which serves different purpose.¹⁸⁷

Nonetheless, although AI legislations across various jurisdictions contain provisions that permit the use of copyrighted materials as training data under specific conditions, it comes with exceptions designed to protect the rights and interests of the copyright holders. These are as follows: (1) option to reserve rights,¹⁸⁸ (2) ensuring fair competition,¹⁸⁹ (3) attribution to copyright holders,¹⁹⁰ (4)

¹⁸⁰ See ARTIFICIAL INTELLIGENCE ACT, *supra* note 84 at § 104.

¹⁸¹ *Id.* at § 105.

¹⁸² § 109.

¹⁸³ See COPYRIGHT ACT OF JAPAN, *supra* note 103 at § 30-4, as amended.

¹⁸⁴ *Id.* at § 30-4 (iii).

¹⁸⁵ § 47-4(1).

¹⁸⁶ § 47-5(1).

¹⁸⁷ See ARTIFICIAL INTELLIGENCE LAW OF THE PEOPLE'S REPUBLIC OF CHINA, *supra* note 107 at § 24.

¹⁸⁸ See ARTIFICIAL INTELLIGENCE ACT, *supra* note 84 at § 105.

¹⁸⁹ *Id.* at § 106.

¹⁹⁰ See DSIT, *supra* note 94 at § 399.

avoiding unreasonable harm,¹⁹¹ (5) incidental and minimal use,¹⁹² and (6) purpose and interference.¹⁹³

The provisions outlined in various AI legislations and policy frameworks reveal a consistent effort to strike a balance between promoting innovation and safeguarding intellectual property rights. While these laws recognize certain permissible uses of copyrighted materials, particularly in the context of AI training, they also impose clearly defined exceptions to protect the interests of the rights holders. These regulatory approaches underscore the importance of responsible AI development.

The implications and legal responsibilities arising from these provisions will be examined in greater detail in the succeeding chapter.

¹⁹¹ See COPYRIGHT ACT OF JAPAN, *supra* note 103 at § 406, as amended.

¹⁹² *Id.* at § 47-4(1).

¹⁹³ See ARTIFICIAL INTELLIGENCE LAW OF THE PEOPLE'S REPUBLIC OF CHINA, *supra* note 107 at § 23.

CHAPTER III

LEGAL RAMIFICATIONS OF COPYRIGHTED MATERIALS IN AI: POSSIBLE LEGAL EXEMPTIONS

Copyright law grants authors with exclusive rights over their original works, including the right to reproduce, distribute, and display their creations.¹⁹⁴ However, copyright law also contains exemptions that allow certain uses of copyrighted materials without the authors' permission. Under the Intellectual Property Code of the Philippines (IPC), one such exemption is the doctrine of fair use, which permits limited use of copyrighted material for purposes such as criticism, commentary, news reporting, teaching, scholarship, or research.¹⁹⁵

In assessing whether the use of a work in a specific instance qualifies as fair use, the IPC of the Philippines provides the following factors to be included in the evaluation: (1) the purpose and nature of the use, including whether it is commercial or non-profit educational, (2) the nature of copyrighted work involved, (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole, and (4) the effect of the use upon the potential market for or value of the copyrighted work.¹⁹⁶

Similarly, the review of various AI legislations across different countries has revealed permissible acts in the use of copyrighted materials in training data for AI applications. These acts include the following:

1. Text and data mining
2. Use of copyrighted data for non-professional and scientific research

¹⁹⁴ INTELLECTUAL PROPERTY CODE OF THE PHILIPPINES, *supra* note 126 at § 177.

¹⁹⁵ *Id.* at § 185.

¹⁹⁶ Filipino Society of Composers, Authors and Publishers, Inc. (FILSCAP) v. Anrey, Inc., G.R. No. 233918, August 09, 2022

3. Use of work for non-personal enjoyment
4. Use of work in computer data processing
5. Use of data for incidental purposes to ensure smooth and efficient operation
6. Use of materials for training data to create new knowledge through computerized data processing
7. Use of copyrighted data for model training with different purposes

In order to gain a clearer understanding of these provisions, it is essential to examine the identified permissible acts alongside its exceptions. This chapter is structured by integrating the findings from the comparative analysis of various AI legislations with insights drawn from the interviews with the participants and relevant academic literature.

Text and Data Mining

Text and data mining (TDM) involves extracting useful information from large datasets which often include copyrighted texts.¹⁹⁷ Exemptions for TDM are essential for AI development, allowing researchers and developers to analyze vast amounts of data without infringing on copyright laws.

The European Union's Directive on Copyright in the Digital Single Market, provides a legal framework that supports TDM for research purposes.¹⁹⁸ The Directive (EU) 2019/790 specifically provides this provision, "*Mandatory exceptions or limitations for uses of text and data mining technologies, illustration for teaching in the digital environment and for preservation of cultural heritage should be*

¹⁹⁷ Thomas Margoni. *Text and Data Mining in Intellectual Property Law: Towards an Autonomous Classification of Computational Legal Methods* (May 16, 2020). CREATE working paper 01/2020 - Forthcoming in Calboli I. & Montagnani L., *Handbook on Intellectual Property Research*, Available at SSRN: <https://ssrn.com/abstract=3602699>

¹⁹⁸ DIRECTIVE (EU) 2019/790, *supra* note 138.

introduced."¹⁹⁹ The existing exceptions and limitations in Union law should continue to apply, including to text and data mining, education, and preservation activities, as long as they do not limit the scope of the mandatory exceptions or limitations provided for in this Directive... This is likewise confirmed by the Director of the Bureau of Copyright and Related Rights in the Intellectual Property of the Philippines, Atty. Emerson Cuyo,²⁰⁰ where he mentioned:

I would say I would be in favor of text and data mining exceptions in the sense that text and data is used to further research. Ganun kasi ang nature ng karamihan sa text and data mining exceptions in copyright laws all over the world. It's for further research to enhance transfer of knowledge, but not really to generate output such as artwork and similar literary work. That one we have to make a distinction. Pag-generative AI siguro, my stand is that they should be compensated because that's use for non-research purposes, although pwedeng non-commercial but we know commercial din yan, yung mga copyrighted works without permission.

I would say yung text and data mining exception is something that we can consider. As for the use of copyrighted works for generative AI training is something that we need to take a second look or to study further because, at the moment, my position is that it's infringing on the copyright of our rights holders, our creators, our authors.

Attorney Cuyo made a categorical distinction between the permissible use of text and data mining as training data to AI applications and generative AI. He expressed support for text and data mining exception when used for research purposes. However, he

¹⁹⁹ *Id.* at § 3 & § 4.

²⁰⁰ Atty. Emerson Cuyo is the incumbent Director of the Bureau of Copyright and Related Rights in the Intellectual Property of the Philippines. The Bureau of Copyright and Related Rights is one of the seven Bureaus in the Intellectual Property Office of the Philippines (IPO). Its main mandate is to implement the state policies on intellectual property and to implement the state policies on copyright and related rights. The Bureau was constituted in 2019, but Atty. Cuyo came only towards the end of 2018. Although it was established or created in 2013 by virtue of Republic Act 10372 that created the Bureau of Copyright and Related Rights, it was actually not until 2018 when the Bureau came was staffed.

opposed granting this exemption to generative AI, as it no longer generates outputs solely for research and the same encroaches upon the rights of the copyright holders, creators, and authors.

Indeed, AI model providers are allowed to use copyrighted materials for text and data mining only if the rights holders have not explicitly reserved their rights.²⁰¹ This means that when rights are reserved, the AI developers must obtain authorization to use the data, except for purposes of scientific and research activities. This exception safeguards the rights holders' control over the use of their works and helps prevent unauthorized exploitation. This is parallel with the case study conducted by Kretschmer, Margoni, and Oruc (2024)²⁰² which states that the European Union's "lawful access" requirement for text and data mining effectively places the decision to permit machine learning in the hands of the rights holders, depending on whether they grant access to their content.

Use of copyrighted data for non-professional and scientific research

Exemptions for the use of copyrighted data for non-professional and scientific research also enable researchers to utilize copyrighted materials to advance knowledge and innovation.²⁰³ This is particularly important in academia where access to a wide range of information is necessary for research and development.

The fair use doctrine of the United States allows the use of copyrighted materials for educational and research purposes under certain conditions.²⁰⁴ This regulation aligns with Atty. Cuyo's views which permits text and data mining solely for research endeavors. This

²⁰¹ ARTIFICIAL INTELLIGENCE ACT, *supra* note 84 at § 104 & § 105.

²⁰² Martin Kretschmer, Thomas Margoni and Pinar Oruc. *Copyright Law and the Lifecycle of Machine Learning Models*. IIC - International Review of Intellectual Property and Competition Law, 55 (2024): 110-138. <https://doi.org/10.2139/ssrn.4670563>.

²⁰³ ARTIFICIAL INTELLIGENCE ACT, *supra* note 84 at § 109.

²⁰⁴ Peter Henderson, Xuechen Li, Dan Jurafsky, Tatsunori Hashimoto, Mark A. Lemley and Percy Liang. *Foundation Models and Fair Use*. ArXiv, abs/2303.15715 (2023). <https://doi.org/10.48550/arXiv.2303.15715>.

exemption is likewise fundamental for fostering academic progress and the development of AI technologies.

In ensuring fair competition, the AI developers must respect the reservations made by rights holders.²⁰⁵ This necessarily implies that any use of copyrighted materials should not create an unfair advantage for the AI developers at the expense of the rights holders. Legislation grounded in this principle seeks to strike a balance between fostering innovation and protecting intellectual property rights. The study conducted by Fernandes (2024) supports this idea, emphasizing the need to balance interests of AI developers and IP rights holders to ensure that technological progress does not come at the cost of creativity and originality.²⁰⁶

Use of work for non-personal enjoyment

AI models that use copyrighted materials for non-personal enjoyment, such as generating art or music, often fall under fair use exemptions.²⁰⁷ These models typically transform the original works in ways that provide new and unique experiences, rather than competing directly with the originals.

Under the Copyright Act of Japan, this form of exploitation is generally viewed as a legitimate use of copyrighted materials, encouraging creativity and development of new forms of expression.²⁰⁸ This finding is consistent with Professor Rebecca Tushnet's²⁰⁹ perspective who expressed that using copyrighted

²⁰⁵ See ARTIFICIAL INTELLIGENCE ACT, *supra* note 84 at § 105 & §106.

²⁰⁶ Pedro Martins Fernandes. *AI Training and Copyright: Should Intellectual Property Law Allow Machines to Learn?* *Bioethica* (2024). <https://doi.org/10.12681/bioeth.39041>.

²⁰⁷ COPYRIGHT ACT OF JAPAN, *supra* note 103 § 30-4.

²⁰⁸ *Id.*

²⁰⁹ Professor Rebecca Tushnet is a professor of law at Harvard Law School. After clerking for Chief Judge Edward R. Becker of the Third Circuit and Associate Justice David H. Souter on the Supreme Court, she practiced intellectual property law at Debevoise & Plimpton before beginning teaching. Her work currently focuses on copyright, trademark and false advertising law. (Harvard Law School, 2022). <https://hls.harvard.edu/faculty/rebecca-tushnet/>

materials as training data is acceptable as long as the AI tool does not produce outputs that are substantially similar to an existing work. She claimed:

If you train on a whole bunch of data, and then it doesn't necessarily produce things that are substantially similar to an existing work, then I think that should be fine. And then I think there's going to be situations where the tool is fine, but it's possible to use it in infringing ways, just like it's possible to use a photocopier or a computer in infringing ways. So, I think there's just a whole lot of different options.

Professor Tushnet added that the use of AI technologies is fine, although there may be instances when it can be used in infringing ways, similar to how photocopiers or a computer are used to infringe the rights of others.

It is, therefore, imperative that when AI-generated outputs are based on copyrighted materials, the AI developers must attribute the original copyright holders.²¹⁰ This condition promotes transparency and acknowledges the contributions of the original creators, thereby protecting their moral and economic rights. The same concern is raised by Eleonora Rosati (2024)²¹¹ in her study where she mentioned that policymakers must give greater attention to issues surrounding AI output generation and use, whether in risk assessment, licensing, or disputes.

Use of work in computer data processing

The use of copyrighted works in computer data processing, such as natural language processing (NLP) or image recognition, is often

²¹⁰ DSIT, *supra* note 94 at § 399. See also Jessica Gillotte. *Copyright Infringement in AI-Generated Artworks* (June 2020). UC Davis Law Review, Vol. 53, No. 5, 2020, Available at SSRN: <https://ssrn.com/abstract=3657423>

²¹¹ Eleonora Rosati. *Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law*. European Journal of Risk Regulation (2024). <https://doi.org/10.1017/err.2024.72>.

considered non-infringing.²¹² This is because the primary goal is not to reproduce the copyrighted work but to use it as a means to achieve technological advancements. This exemption allows AI developers to harness copyrighted data to improve machine learning algorithms and develop more sophisticated models.²¹³ It likewise recognizes the transformative nature of AI development, where copyrighted materials are not reproduced or distributed in its original form, but are instead use to enhance machine learning algorithms.

Furthermore, the use of copyrighted materials is considered reasonable only when it is used in a way that differs from the original purpose for which the work was created.²¹⁴ This means that the use should not replicate the original market function of the work. The use must not likewise interfere with the normal use of copyrighted materials.²¹⁵ This ensures that the original work retains its value and functionality in its intended market.

Use of data for incidental purposes to ensure smooth and efficient operation

The incidental use of copyrighted data to facilitate the smooth and efficient operation of computer systems is another important exemption. This applies to scenarios where copyrighted contents are used temporarily and incidentally, such as caching data for faster access or processing, data indexing, and machine learning model training.²¹⁶ Such uses are generally considered non-infringing as they

²¹² Matthew J. Sag. *Orphan Works as Grist for the Data Mill*. Law & Literature eJournal (2012). <https://doi.org/10.2139/SSRN.2038889>.

²¹³ SAG, *supra* note 212.

²¹⁴ See J. A. T. Silva. *Fair Use in Post-publication Peer Review*. Journal of Educational and Social Research, 5 (2015): 13. <https://doi.org/10.5901/JESR.2015.V5N3P13>. See also S. McIntyre. *Private Rights and Public Wrongs: Fair Use as a Remedy for Private Censorship*. Entrepreneurship & Law eJournal (2012). <https://doi.org/10.2139/SSRN.1998513>.

²¹⁵ See George E. Newman and P. Bloom. *Art and authenticity: the importance of originals in judgments of value*. Journal of experimental psychology. General, 141 3 (2012): 558-69. <https://doi.org/10.1037/a0026035>.

²¹⁶ See ARTIFICIAL INTELLIGENCE ACT, *supra* note 84 at § 105.

²¹⁷ HENDERSON, ET. AL., *supra* note 204.

do not constitute a significant exploitation of the copyrighted material, but only ensures system functionality and efficiency.

Use of materials for training data to create new knowledge through computerized data processing

Using copyrighted data as training material to create new knowledge through computerized data processing is a significant exemption. This practice involves leveraging existing copyrighted works to train machine learning algorithms, which can lead to the development of innovative solutions and insights that were previously unattainable.²¹⁷ This exemption supports the idea that the transformative use of copyright materials to produce new, non-competitive outputs benefits society.²¹⁸

By transforming the original data into new forms of knowledge and utility, this process respects the original intent of copyright while fostering technological growth. It allows AI researchers to push the boundaries of machine learning and artificial intelligence while protecting intellectual property and promoting scientific technological progress.

Use of copyrighted data for model training with different purposes

Use of copyrighted data for model training that serves a different purpose from the original work is often exempt from copyright restrictions.²¹⁹ This acknowledges that AI models typically do not aim to replicate the original content but use it to perform entirely different tasks. For instance, an AI trained on a dataset of books to understand language patterns is not reproducing the books but developing language comprehension capabilities. This transformative use is essential for advancing fields like artificial intelligence and machine

²¹⁸ See Nicolas P Suzor. *Transformative use of copyright material*. (2006).

²¹⁹ ARTIFICIAL INTELLIGENCE LAW OF THE PEOPLE'S REPUBLIC OF CHINA, *supra* note 107 at §24.

learning, where existing data provides the foundation for creating innovative and beneficial technologies.²²⁰ This exemption is crucial for enabling AI to learn from diverse sources without infringing on copyright laws. Hence, the same should be permitted under the copyright law, as it contributes to technological progress without diminishing the value of the original works.

Permissions to use the copyrighted materials, however, do not apply if such use would unreasonably harm the interests of the copyright owners. This includes cases where the use of materials would directly compete with or diminish the market value of the original work.²²¹ The intent is to ensure that the use of copyrighted materials for AI training does not negatively impact the financial or reputational interests of the rights holders.²²²

As discussed in this chapter, the last four exemptions center on the use of copyrighted materials as training data for specific purposes that neither infringe upon the rights of the copyright holders and owners, nor diminish the value of their original works. These exemptions recognize that utilizing copyrighted content for machine learning and artificial intelligence training can lead to the creation of new, transformative technologies and applications. By ensuring that the use of such data is limited to purposes that do not replicate or distribute the original material, these provisions strike the balance between fostering innovation and respecting intellectual property rights. Consequently, this approach supports technological progress while safeguarding the interests and economic value of the original creators' works.

²²⁰ Royani, Royani, Sondang Deri Maulina, Sugiyono Sugiyono, Rio Wahyudin Anugrah, and Brigitta Callula. 2024. *Recent Developments in Healthcare Through Machine Learning and Artificial Intelligence*. IAIC Transactions on Sustainable Digital Innovation (ITSDI) 6 (1):86-94. <https://doi.org/10.34306/itsdi.v6i1.680>.

²²¹ See I. Polishchuk and V. Koloniuk. *To The Question Of Defining The Concept Of Forging Art Works*. (2020): 672-684. <https://doi.org/10.33994/kndise.2020.65.66>.

²²² POLISHCHUK & KOLONIUK, *supra* note 221.

This perspective aligns with the views of the participants of the study who support the use of copyrighted materials as training data for AI applications, provided certain conditions are met and the provisions of copyright law are respected.

In particular, Professor Tushnet highlighted that these conditions could be likened to the broad exemptions found in the fair use provision of the US Copyright Law which parallels the fair use provisions of the Intellectual Property Code of the Philippines (IPC). She emphasized that such exemptions are crucial for balancing the promotion of technological innovation with the protection of creators' rights. By ensuring that the use of copyrighted materials for AI training is both purposeful and transformative, these provisions can facilitate significant advancements in AI while maintaining the integrity and value of the original works. She stated:

[AI programs should] not be exempted, but also not especially targeted. I think our regular copyright rules are probably good enough... [In it,] you have a right to prevent the copying of your works unless an exception applies, including fair use and...I think that's a workable regime. I think fair use can probably handle AI one way or the other, so I don't think we need special rules for it. We already have rules about copying in general, and we already have defenses for particular kinds of uses, including fair use.

In the U.S. we tend to do everything under fair use. But since many other countries don't have a broad fair use exception... a number of them decided to enact specific machine learning exemptions. So, I think having a flexible fair use provision instead of a closed list of exceptions, which is the case in many other countries, is much better.

Professor Tushnet believes that fair use provision can effectively address the unique challenges posed by AI, negating the need for special regulations. She contrasted this viewpoint with the practices in other countries that lack a broad fair use provision and have opted for specific machine learning exemptions. The flexibility of the US fair use provisions is deemed as advantageous compared to the more rigid, closed list of exceptions found in other jurisdictions. She underscored

that the flexible legal framework like fair use can adapt to evolving technologies such as AI, providing necessary protections without stifling innovation.

This idea is supported by Atty. Cuyo who earlier advocated the inclusion of the specific provision in copyright law concerning the use of copyrighted materials in text and data mining. However, he suggested that rather than merging it with the fair use provisions of the IPC, it would be more effective to include a distinct provision for text and data mining under Article 184 of the IPC which provides for specific limitations in the copyright. He explained:

Actually, when we were discussing internally among us in the Bureau where a possible text and data mining exception can come in. In particular, [we have identified] Section 184.1 of the Code where you have the various situations that are considered as exceptions and limitations. Medyo mahirap siyang ipasok under Section 185 which is the General Fair Use Provision, yung four factors test ...because pag doon mo siya ipasok it will limit actually the flexibility of the General Fair Use Provision.

Yan nga yung strong feature ng General Fair Use Provision that it can actually adapt to the times, it can adapt to technology for as long as those four factors or four requirements are not breached. So, I think the best provision to put it in is under the present Section 184.1 of the Code.

Atty. Cuyo argued that integrating the proposed provision into the fair use framework might constrain its flexibility, which is crucial for adapting to new technologies and evolving practices. By establishing a dedicated provision, the law can clearly address the unique aspects and requirements of AI-driven data mining while preserving the adaptive strengths of the general fair use provision. This approach aims to provide legal clarity and support innovation without undermining the broad applicability and adaptability of fair use.

Notably, Atty. Darwin P. Angeles,²²³ concurred with Professor Tushnet on utilizing the fair use provision to recognize the use of copyrighted materials in training data as an exception. He also agreed with Atty. Cuyo, suggesting that distinguishing the use of copyrighted works for training data might be more straightforward than presumed. He argued:

Fair use kasi is contextualized based on the purpose, commercial versus non-commercial. Non-commercial sometimes can go into the realm of art... I would say the artist being commissioned to do work and drawing inspiration from copyrighted works and using AI in that regard can invoke a stronger fair use case than the other one purely commercial.

[In terms of using works as training data], I think they're easy to draw. For instance, if the database is published online [which] has a certain licensed to use, it's available to the public, and it grants a license to use for a wide variety of purposes... Technically may copyright yung creation of the digital image but the subsisting work is actually based upon part of the public domain na e. So, if you use those images... I don't think that would be too contentious na to think may copyright on that. The photograph of Monalisa for example, is one of the most common derivative work. But moral rights kasi are forever... Pero may mga practical considerations rin na is it even worth doing that or you have to be strict in those instances. So, a carve-out can be made pero hindi siya pwedeng too broad or too narrow.

Atty. Angeles also contended that the flexibility inherent in the fair use doctrine allows it to adapt to various contexts, including AI training. By applying the fair use factors, it is possible to assess the purpose, nature, and impact of the use, ensuring that non-commercial,

²²³ Atty. Darwin Angeles is a Partner and the Head of the Intellectual Property (IP) Department of the MOSVELDTT Law Offices. His practice also covers dispute resolution and data privacy tech law. He is likewise listed as a notable practitioner in Dispute Resolution by the Legal 500 Asia-Pacific. Atty. Angeles specializes in patent prosecution for computer-implemented technologies, commercialization of IP assets, and IP rights enforcement. He has registered multiple patents involving blockchain, data science, and artificial intelligence technologies. (MOSVELDTT Law Offices, 2019). <https://mosvelddttlaw.com/peoples/angeles/>

transformative uses, such as AI training, are adequately protected. This approach maintains the adaptability of fair use while still respecting the rights of copyright holders, thus providing a balanced and practical solution to challenges posed by AI technologies. However, in terms of training data, he asserted that there might be a need for a specific carve-out in copyright law, something that is not too broad nor too narrow. This carve-out would ensure that AI development can progress without overly restrictive limitations while still protecting the interests of copyrights holders. Striking the right balance in such legislation would be crucial to fostering innovation and maintaining the integrity of intellectual property rights.

Exemptions in the use of copyrighted materials for AI training are essential for balancing the protection of intellectual property with the need for technological advancement. These exemptions enable researchers and developers to utilize copyrighted data in ways that promote innovation, creativity, and scientific progress without infringing on the rights of copyrights holders. As AI technology continues to evolve, it is crucial to maintain and adapt these exemptions to ensure they effectively support the dynamic and transformative nature of AI research and development.

CHAPTER IV

BETWEEN CODE AND CREATIVITY: LEGAL RESPONSIBILITIES OF AI ACTORS FOR THE UNAUTHORIZED USES OF COPYRIGHTED MATERIALS

Copyright infringement claims could arise in two neural network-based machine learning scenarios. First, infringement may occur when an engineer compiles digital dataset for training purposes where data selection of copyrighted works is digitized and/or produced without authorization from the copyright holder.²²⁴ Second, copyright infringement claim may likewise occur when unauthorized intermediate copies of original works are made during training in neural networks.²²⁵ Hence, in order to determine the responsibilities of AI actors in the unauthorized uses of copyrighted materials, the researcher adopted the Generative AI Supply Chain developed by Lee, Cooper, and Grimmelmann (2023)²²⁶ as a framework in discussing copyright infringement on the different stages of AI process.

Furthermore, when dealing with AI and copyright issues, it is crucial to distinguish between the use of the copyrighted works during the “*AI development or training stage*” and the infringement occurring at the “*generation or utilization stage*.”

On one hand, the *AI development or training stage* involves the creation of training dataset by gathering and reproducing copyrighted works for use as training data, and developing AI (pre-trained machine learning models) using these datasets. On the other hand, the *generation or utilization stage* includes generating materials (such as images, texts, etc.) using AI, uploading and publicly sharing AI generated images, and selling copies of these generated images (e.g.

²²⁴ GILLOTTE, *supra* note 36.

²²⁵ *Id.*

²²⁶ Katherine Lee, Feder Cooper, & James Grimmelmann. *Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain* (July 27, 2023). Forthcoming, *Journal of the Copyright Society* 2024, Available at SSRN: <https://ssrn.com/abstract=4523551> or <http://dx.doi.org/10.2139/ssrn.4523551>

publishing art collection books). In simple terms, the former pertains to the input while the latter focuses on the output.

Copyright Infringement on the Different Stages of AI process

The input side encompasses more intricate stages involving multiple actors with different roles and liabilities. According to Lee, Cooper, and Grimmelmann (2023),²²⁷ these stages include (1) the creation of expressive works, (2) data creation, (3) dataset collection and curation, (4) model pre-training, (5) model fine-tuning, (5) system deployment, (6) generation, and (7) model alignment.

On the **creation of the expressive works**, the works of the artists and writers were transformed into data for machine learning.²²⁸ This includes transforming physical works into digital copies that can be uploaded online. Once available online, these creations become accessible for AI development, potentially implicating copyright issues if used without authorization.

This is the reason for the inclusion of the *transparency and data disclosure* as one of the responsibilities of the AI actors.²²⁹ The GPAI models are required to produce and publish a detailed summary of their training data. The summary must comply with the copyright laws, ensuring that the copyright holders can understand and enforce their rights.²³⁰ This transparency helps prevent the misuse of copyrighted materials and fosters trust between AI developers and copyright holders. AI developers must also publicly report on the limitations, capabilities, and appropriate uses of their AI systems.²³¹ This includes disclosures about the data used for training models and details on data processing and testing. Such measures ensure that the

²²⁷ LEE, COOPER, AND GRIMMELMANN, *supra* note 226.

²²⁸ *Id.*

²²⁹ See ARTIFICIAL INTELLIGENCE ACT, *supra* note 84 at § 105.

²³⁰ See Andres Guadamuz. *The EU's Artificial Intelligence Act and Copyright* (July 01, 2024). *Journal of World Intellectual Property*, Forthcoming <http://dx.doi.org/10.2139/ssrn.4997854>

²³¹ *Id.*

AI development process is visible and accountable, preventing potential copyright infringement.

The primary responsibility in this stage lies with the developers who used said works for training AI systems without proper permissions. It is worth noting that conversion of the original works into digital formats that AI systems use retains the copyright, posing potential infringement risks.

The second stage, **data creation**, pertains to the extensive use of datasets for training and validation.²³² In this stage, the original expressive works are different from their *datafied* counterparts (digital copies or creative works that were transformed into computer-readable formats).²³³ In this process, the quality and quantity of data significantly impacts the authenticity and refinement of AI outputs. Unfortunately, the selection of materials for datasets often includes copyrighted materials acquired without authorization. Hence, the liability primarily falls on data collectors and curators who compile the datasets since they must ensure that the data used complies with copyright laws.

The data collected may be further filtered out through curation or the process of filtering out undesirable contents based on the model created (*e.g.* toxic speech).²³⁴ At this stage, **dataset creation and curation** often occur simultaneously which may be performed by the same parties who are responsible for training generative AI models.

In both stages of data creation and dataset curation, *compliance with the copyright law* is imperative. The provisions of the AI Act specifically state that developers of the GPAI models must adhere to the Union Copyright Law when publishing summaries of their

²³² LEE, COOPER, AND GRIMMELMANN, *supra* note 226.

²³³ Rita Matulionyte. *Australian Copyright Law Impedes the Development of Artificial Intelligence: What Are the Options?* (October 27, 2020). Available at SSRN: <http://dx.doi.org/10.2139/ssrn.3720289>

²³⁴ A. Borji. *A Categorical Archive of ChatGPT Failures*. ArXiv, abs/2302.03494 (2023). <https://doi.org/10.48550/arXiv.2302.03494>.

training data.²³⁵ Compliance therein is crucial for maintaining the legality of their operations and respecting the rights of copyright holders. The UK government likewise emphasizes trust and transparency among all parties, particularly regarding how AI developers use copyrighted data and attribute outputs to copyright holders.²³⁶ This stance ensures that developers are held accountable for their use of copyrighted materials. Nonetheless, in this stage, there is a potential copyright liability among data creators and curators independent from the potential liability of model trainers.²³⁷

The next stage is **model training** which requires vast volumes of data, often scraped from the internet, including both copyrighted and non-copyrighted works.²³⁸ In the data collection and curation process, the data dataset creators filter out undesirable content based on the model's needs. They have the responsibility to ensure that the data they collect, and curate does not infringe on copyrights, otherwise, it can lead to legal repercussions.

In the **model pre-training stage**, the model trainers select a dataset, model architecture, training algorithm, and seed value for training.²³⁹ This iterative process involves human intervention and results in a pre-trained model or base model. The responsibility at this stage lies with the model trainers who must ensure that the training process adheres to legal standards and does not use unauthorized copyrighted materials. This liability can further extend to entities that later use the model for additional training.

The fifth stage, **model release and system deployment**, allows the AI models to become accessible for others to use or modify.²⁴⁰ The liability in this stage can be associated with the entity who released the

²³⁵ ARTIFICIAL INTELLIGENCE ACT, *supra* note 84 at § 108.

²³⁶ DIRECTIVE (EU) 2019/790, *supra* note 138 at § 31.

²³⁷ LEE, COOPER, AND GRIMMELMANN, *supra* note 226 at 38.

²³⁸ *Id.*

²³⁹ *Id.*

²⁴⁰ *Id.*

model, ensuring that the same does not facilitate copyright infringement. They must also provide clear guidelines on how the model can be used without violating the copyright laws.

In both stages, AI actors are must *ensure the quality of their training data*. AI developers may also be held responsible for improving the quality of their training data by ensuring its veracity, accuracy, objectivity, and diversity.²⁴¹ They must also take appropriate action if they discover any prohibited information in the training data. High quality data were recognized as essential in creating reliable AI models and minimizing legal risks associated with using copyrighted materials.²⁴² Businesses are likewise expected to handle copyrighted issues pragmatically, ensuring fair treatment and clarity. This approach helps mitigate conflicts and promotes a fair use of copyrighted materials within the legal framework.

In the **generation stage**, users have various methods to initiate the generation process, most common of which is the use of prompt to produce outputs.²⁴³ If a user wants to produce generations using the released model, they must develop a code that could interact with the model parameters to execute the generation process.²⁴⁴

Nonetheless, majority of users interact with the models indirectly through a service provided by model deployer like an application programming interface (API) or a web application.²⁴⁵ In such a case, a user typically selects an application (*e.g.* ChatGPT, Midjourney, DALL-E, etc.) with the certain outputs in mind, ensuring that their selection aligns with their intentions which may relate to concerns about potential infringement.²⁴⁶ In this stage, understanding the relationship between the user and the selected deployed system is

²⁴¹ HOUSE OF REPRESENTATIVES 6212, *supra* note 76 at § 301 (a)(1)(C).

²⁴² ARTIFICIAL INTELLIGENCE LAW OF THE PEOPLE'S REPUBLIC OF CHINA, *supra* note 107 at §45.

²⁴³ LEE, COOPER, AND GRIMMELMANN, *supra* note 226.

²⁴⁴ *Id.* at 49.

²⁴⁵ *Id.* at 50.

²⁴⁶ *Id.*

a crucial aspect of analyzing copyright infringement. This relationship is characterized in three distinct viewpoints.

First, the user actively uses prompts to generate outputs to which the application responds passively. On this view, the user is a potential direct infringer while the application functions akin to a neutral technological provider like the web host or the internet service provider (ISP).²⁴⁷ Second, the system takes an active role in output generation while the user remains passive.²⁴⁸ In this perspective, the user can be likened to a viewer of an infringing broadcast or an unwitting purchaser of a pirated book. The primary copyright responsibility rests with the deployed systems, potentially extending to other entities in the generative supply chain.²⁴⁹ Third, both the user and the system actively contribute to generating infringing outputs. In this context, the user is like a patron who commissions a copy of a painting while the system acts as the artist executing the request, the share is a goal of producing an infringing work.²⁵⁰

In this regard, *accountability measures* must be exercised. In doing so, specific roles dedicated to AI safety must also be assigned, and training for developers and deployers of AI products is mandated in certain settings. These measures ensure that individuals involved in AI development understand their responsibilities and the legal implications of using copyrighted materials.

Lastly, in the **model alignment stage**, the overarching aim of model alignment is to align model outputs with specific generation preferences, the most popular alignment technique is called reinforcement learning with human feedback (RLHF). As the name suggests, RLHF combines collected human feedback data with a

²⁴⁷ Disha Shrivastava, H. Larochelle, and Daniel Tarlow. *Repository-Level Prompt Generation for Large Language Models of Code*. (2022): 31693-31715. <https://doi.org/10.48550/arXiv.2206.12839>.

²⁴⁸ SHRIVASTAVA, LAROCHELLE & TARLOW, *supra* note 247.

²⁴⁹ *Id.*

²⁵⁰ *Id.*

(reinforcement learning) algorithm in order to update the model.²⁵¹ Human feedback data can take a variety of forms, which include user ratings of generations.²⁵²

Future training and alignment on the model may include both the input prompt and the generation in addition to the feedback provided.²⁵³ User-supplied prompts may include copyrighted content created by either the use themselves or by another party.²⁵⁴

During the early stages of training, when a relatively low amount of content is fed into the system, the output produced might be quite similar to the input data or can contain substantial parts from work inputted.²⁵⁵ In these and other cases, the output that contains entire or substantial parts of the works that were fed into the system is likely to infringe the reproduction right associated with pre-existing copyrighted subject matter.²⁵⁶ Additionally, if an infringing output is further communicated to the public, this may trigger other exclusive rights, such as the public communication right, or publishing right.²⁵⁷

The Responsibilities of AI Actors in the Unauthorized Uses of Copyrighted Materials

The responsibility and liability of AI actors vary across different stages of the AI process. Developers, data collectors, model trainers, and users each have distinct roles and responsibilities in ensuring compliance with the copyright laws.

²⁵¹ Zheng Yuan, Hongyi Yuan, Chuanqi Tan, Wei Wang, Songfang Huang and Feiran Huang. *RRHF: Rank Responses to Align Language Models with Human Feedback without tears*. ArXiv, abs/2304.05302 (2023). <https://doi.org/10.48550/arXiv.2304.05302>.

²⁵² YUAN, ET. AL., *supra* note 251.

²⁵³ LEE, COOPER, AND GRIMMELMANN, *supra* note 226.

²⁵⁴ *Id.* at 50.

²⁵⁵ *Id.*

²⁵⁶ *Id.*

²⁵⁷ *Id.*

In relation to the different stages in the AI process, it can be inferred that the primary responsibility among the AI actors rests with the developers as clearly demonstrated in stages one through five while users assume shared responsibility with AI developers only in stages six and seven. It is worth highlighting that while data collectors and model trainers may be identified separately within the process, they are typically affiliated with and employed by an entity or corporation involved in AI development who has the responsibility over the acts committed by its employees. This observation resonates with the views shared by the Director of the Bureau of Copyright and Related Rights in the IPOPHL, Atty. Emerson Cuyo, who stated that:

I think the responsibility to acknowledge, to attribute, to compensate, to request permission from the copyright owners or the rights holders of copyrighted materials used for AI development, rest with the developer. Usually, in the model that is coming out now, it's usually an entity. I haven't heard of many small players or individuals doing AI development.

Usually, it's an entity, many of them large corporations. So, I think the responsibility rests with them to be able to make sure that in training their AI models, the appropriate permission is used, the proper permission is used, or in case there is already a provision or standard under the law, then they should be compensated. So, the accountability for compensating and for seeking permission...at this point is seeking permission upon use, the compensation must be properly paid to the copyright owner, rest with ...the developer.

The same views aligned with the opinions expressed by Atty. Darwin P. Angeles, a private law practitioner in the Philippines specializing in the intellectual property. In an interview with the researcher, he mentioned that the party accountable for the damages arising from the unauthorized uses of the copyrighted materials to AI models varies depending on the specific circumstances of the case. He has given multiple examples when an AI developer may be held liable for the damages suffered from the unauthorized uses of the copyrighted materials:

It depends on the kind of case... kasi kung after the fact naman, kunwari you have a company, and they engaged an AI company as a service tapos gumawa ng work product, tapos ginamit yung work product...shouldn't the operator also be made liable?... Or let's take it differently, kunwari, AI is being used din for software development, video game development... [and] the AI creates a work product based on copyrighted material, tapos no appropriate license was obtained, therefore.

The problem is when you go to commercial use, for example, ako I'm the lawyer for the publisher, I would demand that before we ink that contract, there must be a warranty of a good title on the part of the training data ng AI so that covered yung company ko from claims of damages. In fact, I would even go to what is called as the "harmless clause." So, if my client – for example – Electronic Arts gets sued for the use of open AI software in theory, I can pass the buck. I can claim that "sabi nyo, clean yung data nyo, bakit ako kinakasuhan ito? Harapin nyo sila."

So, between the developer and the client, yung developer [yung liable] kasi you offer AI as a service, you have to make sure na pag ginamit ko yan hindi ako sasabit. Bakit pa kita babayaran? E di sana nag-free service na lang ako. Tapos mag-caveat emptor ako. If you want my money for your AI, you have to deliver me a good product.

This entails that developers must verify that the data used in training AI must have already complied with the provisions of the copyright law as early as the scraping stage of the data collection, otherwise, their AI products may expose their clients to potential lawsuits.

The case of *Roberto Mata v. Avianca, Inc.*²⁵⁸ is demonstrative in this case. In the said case, Mata's attorney, Steven Schwartz, was sued for using ChatGPT in his legal brief after the court found out that several of his cited cases were fictitious, non-existent, or grossly misrepresented.²⁵⁹ The court further stated that the names, citations,

²⁵⁸ Roberto Mata v. Avianca, Inc., 22-cv-1461 (PKC), U.S. District Court for the Southern District of New York, June 22, 2023.

²⁵⁹ MATA V. AVIANCA, INC., *supra* note 258.

and content of the cases provided in his brief did not correspond to any actual legal precedent.²⁶⁰

This is likewise affirmed by the case of *hiQ Labs, Inc. v. LinkedIn Corp.*²⁶¹ where LinkedIn was sued for scraping information from public LinkedIn profiles to provide insights and analytics on employee skills to employers. In this case, the Court ruled in favor of LinkedIn stating that the company should not be penalized for accessing publicly available data, as these data were not protected by the authentication barriers.²⁶² The same is distinguished from unauthorized access to protected data which are subjected to a certain restriction under the law.²⁶³

On the level of the users, most of the literature and other cited authorities emphasized that users share responsibility with AI developers in cases where damages are incurred in an unauthorized use of copyrighted materials to AI models. However, the participants disagree with this view, arguing that direct liability should apply for direct infringement, as vicarious or contributory infringement is not punishable under our jurisdiction. Atty. Cuyo explained:

[There] are nuanced situations that puts the user in liability or potential liability for copyright infringement. Kasi there's direct infringement. Right now, I think our present law, does not even punish the users unless...[they] are a direct infringer meaning ikaw mismo yung nag-reproduce, nag-publicly communicate.

But if you're asking me, 'pag nag-prompt ka lang probably of something that's already there na, you didn't do a direct act to infringe on the right of another, I think medyo far-fetched ang liability ng user for copyright infringement because, I think, that should be the responsibility of AI development company...if it is something there, and you do not do a direct infringing act, then, I

²⁶⁰ *Id.*

²⁶¹ *hiQ Labs, Inc. v. LinkedIn Corp.*, 938 F.3d 985 (9th Cir. 2019)

²⁶² *HIQ LABS, INC. V. LINKEDIN CORP.*, *supra* note 261.

²⁶³ *Id.*

think, the chances that you'll be held accountable is low or none at all.

This narrative is affirmed by Atty. Angeles who mentioned that “for the training data cases, [the liability] should lie with the person who caused the unlawful use of, and the production of the copyrighted data for purposes of training the AI.” This indicates that in cases involving training data, the responsibility lies with the individual who instigated the unlawful usage and creation of the copyrighted materials for training purposes of AI. This scenario is akin to the case of *ABS-CBN v. Gozon, et.al*²⁶⁴ where the respondents admitted to using ABS-CBN’s copyrighted materials in the live feed of GMA’s TV program, without securing proper authorization or licensing from ABS-CBN. The Court ruled that the respondents’ unauthorized use of ABS-CBN’s copyrighted materials constituted a violation of its exclusive rights as the copyright owner.²⁶⁵ By incorporating ABS-CBN’s copyrighted materials into their own work without permission, they infringed upon ABS-CBN’s rights to control the reproduction and distribution of its copyrighted content.²⁶⁶

The Court likewise ruled that the “officers and employees of GMA cannot be held liable for infringement under the Intellectual Property Code since it does not expressly provide direct liability of the corporate officers.”²⁶⁷ Instead, the persons directly responsible for the unlawful reproduction and use of copyrighted materials must be held liable for copyright infringement.

In contrast, Professor Rebecca Tushnet expressed that liability of an actor can only occur if the resulting output produced by the AI tool is substantially similar to an existing protected work, but no liability should attach if the output is not substantially similar, even if

²⁶⁴ ABS-CBN Corporation v. Felipe Gozon, et. al., G.R. No. 195956, March 11, 2015

²⁶⁵ ABS-CBN CORPORATION V. GOZON, ET. AL., *supra* note 264.

²⁶⁶ *Id.*

²⁶⁷ *Id.*

copyrighted materials were used in the training data. She explained that:

[There is] a pretty significant difference between training a model only on one author's works and then asking for new works that continue the story told in the previous works. That is, I think, is at least possibly going to infringe. On the other hand, if you train on a whole bunch of data, and then it doesn't necessarily produce things that are substantially similar to an existing work, then I think that should be fine. And then there's going to be situations where the tool is fine, but it's possible to use it in infringing ways, just like it's possible to use a photocopier or a computer in infringing ways. So, I think there's just a whole lot of different options.

When the AI applications already produced a particular output... [that] is substantially similar to probably an existing artwork or an existing paper, that is where infringement would occur, assuming that it didn't come from unprotectable similarities. So, there might be cases where, especially if you're asking for a description of a historical factual event, they might end up being very similar. So, you would need to know, was that copied into the training set and was it memorized by the algorithm? Because if not, maybe the similarity is just due to similarity of unprotectable expression. But it's certainly possible that, that could infringe...So, I think in the US, a variety of kinds of training are likely to be fair use, [hence] I don't think you can answer that question in a blanket way. But I think there are a bunch of potential uses that would not be infringing.

This means that the fair use provision under the Copyright Law of the United States²⁶⁸ can serve as a defense in copyright infringement cases involving the unauthorized uses of copyrighted materials to AI models. The fair use provision of the said law allows the use of copyrighted works for purposes like criticism, teaching, or research without being considered copyright infringement.²⁶⁹

²⁶⁸ U.S. Copyright Office, *Compendium of U.S. Copyright Office Practices* § 107 (3d ed. 2021).

²⁶⁹ COMPENDIUM (THIRD), *supra* note 268.

The US Copyright Law provides four specific factors that needed to be considered in deciding whether the case qualifies for fair use. These factors include (1) the purpose and character of the use, including whether such use is of a commercial nature or for nonprofit educational purposes, (2) the nature of the copyrighted work, (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole, and (4) the effect of the use upon the potential market for or value of the copyrighted work.²⁷⁰ It further provides that the “fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors.”²⁷¹ Various legal precedents in the US have likewise increasingly recognized the transformative use in fair use determination. In cases like *Authors Guild v. Google, Inc.*²⁷² the courts have acknowledged that transformative uses, even if commercial, can qualify as fair use.

Interestingly, the fair use provision in the Philippine copyright law mirrors that of the US. This similarity suggests that the arguments and defenses based on fair use in the US could be persuasive or applicable in the Philippines as well. This transnational consistency in copyright principles enhances the viability of fair use defenses for AI developers operating in both jurisdictions.

The exploration of liability concerning AI actors in copyright matters reveals a dual responsibility existing across both the training (input) and utilization (output) stages of AI models. Notably, the examination of its input side unveils a clearer definition of liability for AI developers and users, contingent upon their actions and participation in each phase of the AI process. This assertion finds support in the narratives of the participants who provided specific examples of infringing acts, drawing from both the theoretical understanding and practical experiences within their profession. Additionally, the integration of the case laws as references serves to

²⁷⁰ *Id.* at § 108.

²⁷¹ *Id.*

²⁷² *Authors Guild v. Google, Inc.*, No. 13-4829 (2d Cir. 2015)

bolster key arguments and elucidate specific points which further reinforced the understanding of AI actors' liability in copyright contexts.

In this view, it can be concluded that the primary responsibility among AI actors in the unauthorized use of copyrighted materials lies with the developers, including the data collectors/curators, and model trainers. Direct liability may also be attributed to users who committed direct infringing acts in utilizing AI applications. This underscores the importance for developers to ensure that the data used in training AI adheres to copyright law provisions, starting from the scraping stage of data collection, to mitigate potential legal risks in using AI products. Furthermore, the invocation of the fair use provision emerges as a pivotal defense mechanism against the identified liabilities, underscoring the multifaceted nature of copyright issues in AI.

CHAPTER V

SAFEGUARDING INNOVATION: UPDATING THE PHILIPPINE IP CODE FOR THE AGE OF AI

Albeit international law provisions permit the use of copyrighted materials as training data, there remains a crucial need to safeguard the rights of copyright holders by identifying limitations to these permissible acts and ensuring that these limitations are properly met. Coupled with the permissible acts and their limiting conditions are the responsibilities and liabilities of AI developers in copyright-related concerns, particularly concerning the use of copyrighted materials in training AI models. This includes transparency and data disclosure, compliance with copyright laws, maintaining data quality and integrity, and accountability measures.

Nonetheless, the challenge with these provisions lies in effectively enforcing the rights of copyright holders and ensuring that AI developers comply with the legal requirements designed to protect these rights. Achieving compliance necessitates robust enforcement mechanisms and regular audits to verify adherence to legal standards. In relation to this, the participants of the study have expressed their views on regulating the use of copyrighted materials and have proposed recommendations to ensure that the rights of copyright holders are protected.

As to regulation, Atty. Cuyo circled back to the accountability of AI developers in the unauthorized uses of copyrighted materials as training data for AI models. He stressed that since responsibilities and accountabilities rest upon AI developers, they needed to be regulated to ensure that owners and copyright holders are compensated. Attorney Angeles concurred with Atty. Cuyo on the need for regulations, but expressed reservations about their implementation, noting that global consensus on the issue has yet to be reached. He answered when asked about the need for regulations on copyright, *“I think so, pero mahirap i-cross yung bridge na yan because... courts from different sides of the world are treating the copyright question differently, so*

it might benefit us to kick the can down the road so to speak and watch which one is the more effective one." This implies that given the lack of global consensus on copyright issues, it would be prudent for the Philippines to observe first and assess various methods and approaches to determine which might be the most effective, before adopting same principles and implementing similar regulations.

Conversely, Prof. Tushnet disagreed with the other participants, contending that the main challenge regarding copyright is not compensation but the preservation of human creativity in a world increasingly dominated by AI.

I think that copyright should not be the barrier. And one thing that's very evident in what is happening now with big companies cutting licensing deals with specific companies is the question, are there going to be huge data sets? The question is, who's going to get paid for them? And so, from the perspective of the artists of five years from now, the ones who aren't making art yet, the AI is coming for them. It's going to be competing with them, even though it didn't copy any of their work.

So, we should be thinking about the challenges, not as a question of who's going to get paid for creating the training set, but how are we as a society going to react to the output? And how are we going to preserve human creativity in that world? So, I don't think that mandating that copyright owners get payment now is very helpful. It's just about dividing up money right now, but it's not a solution for the long term.

Professor Tushnet raised important questions about the evolving landscape of copyright in the age of AI. She emphasized that with the major corporations engaging in licensing deals for access to datasets, the concern arises not only about compensation for the creators of training datasets but also about the impact of AI-generated content on future artists and creatives. She asserted that the looming presence of AI in the creative sphere suggests a need to reevaluate our approach to copyright and consider broader societal implications beyond immediate financial compensation. Instead of simply dividing revenue among stakeholders, there is a call to address the larger

challenge of preserving human creativity in a world increasingly influenced by AI-generated content. This viewpoint suggests that focusing solely on payment for copyright owners may not offer a sustainable solution in the long run and urges a deeper exploration of how society can navigate the complexities of AI-driven creativity while safeguarding the integrity of human expression.

As to recommendations to further improve the copyright law, Prof. Tushnet emphasized that there is no need to amend the US Copyright Law since its fair use provision is already broad and encompassing to address issues on the use of copyrighted materials as training data in AI models. She pointed out that it is “capacious” to recognize training data as fair use when appropriate. Attorney Angeles affirmed her views, asserting that there is no need to amend the Philippine Copyright Laws to address use of copyrighted materials in AI training. Instead, he noted that the same can be done by the IPO as the government agency vested with the administrative and adjudicatory powers necessary to manage the repository of copyrighted materials. Nonetheless, even if he does not support the integration of AI-specific provisions into the existing copyright law, he recommended measures to protect copyright through the use of digital rights signatures on digitized materials to enable users to trace the origins of certain materials.

The perspectives of Prof. Tushnet and Atty. Angeles align with the views of Atty. Cuyo, who focused on the importance of practical enforcement of the IPC to safeguard copyrights. He pointed out that while copyright law stipulates the protection afforded to an individual copyright, there remains a need for practical measures to ensure that these protections are effectively enforced. Specifically, he suggested the following measures: (1) valuation of copyright holders’ work, (2) disclosure requirements, (3) registration of datasets with IPO, (4) fair learning option, and (5) digital receipts via blockchain.

First, Atty. Cuyo underlined that he considers AI as a tool which can acquire permanence in our technological landscape, necessitating the incorporation of standards into copyright law to address the use of

copyrighted materials in AI training. One of its essential aspects is the **valuation of creative works** where he expressed:

I look at AI kasi as a tool. So, since AI is here to stay, we should provide standards in the law,... all of those are improvements in the IP code that must be in order to catch up with the trend in AI development. I think ang important aspect... na makonsider yung valuation nila because just to put a premium on human generated works I think in terms of valuation may premium 'pag human if it becomes inevitable actually to stop AI generated outputs... I think we should be careful dun sa valuation part to give still premium to humans... [as compared to the] mass produced... AI generated works, kapag human baka kailangan mong bigyan ng premium.

Attorney Cuyo mentioned that human-generated works should be given a premium valuation to acknowledge their unique creativity and effort, distinguishing them from mass-produced AI-generated outputs. This premium not only incentivizes human creativity but also respects the intrinsic value that human authors bring to their work. By carefully considering these valuation standards, we can ensure that human creators continue to receive the recognition and compensation they deserve, even as AI becomes increasingly integrated into the creative processes. This balanced approach will help protect the interests of human creators while accommodating advancement and capabilities of AI technology.

Second, as to **disclosure requirements**, Atty. Cuyo put emphasis on the accountabilities of AI developers and development companies as primary actors who will be held liable for unauthorized uses of copyrighted materials as training data for AI applications. He mentioned that this necessitates regulatory measures that can take various forms including the mandatory disclosure of datasets used in training AI models. By requiring transparency in dataset disclosure, developers would need to demonstrate that the data used complies with existing domestic and international copyright laws. He explained:

Siguro, babalik ako doon sa napag-usapan natin kanina that the accountability rests with the AI developers or AI development company. So, the use of copyrighted data as input is their accountability. So yes, they should be regulated, yes, they should be made to compensate the use of copyrighted materials for training their AI models.

Yung algorithm siguro mahirap because as you said that may fall within the ambit of trade secrets na yan. But yung training dataset nila, I think, we can actually require them to disclose.

Pero how they are going to comply with it, I'm trying to come to terms with how we can actually implement yung transparency na tinatawag nila. So yes, they should be still regulated in terms of disclosure and transparency but how do they comply with it is something that hindi ko pa maisip sa ngayon.

Attorney Cuyo likewise mentioned that AI developers should be obligated to compensate rights holders for the use of their copyrighted materials. The financial restitution acknowledges the value of the original works and ensures that creator receives fair compensation for their contributions. He also pointed out that while regulating algorithms might be challenging due to their classification as trade secrets, the transparency of training datasets is a feasible area for regulation.

Implementing these transparency requirements poses a challenge, as the specific mechanisms for compliance need careful consideration. Developing a practical framework for disclosure and verification is essential to ensure that AI developers adhere to these standards without compromising their proprietary information. Ultimately, the goal is to balance the need for innovation in AI with the protection of intellectual property rights, ensuring that the benefits of AI advancements do not come at the expense of creators' rights and compensation.

Third is the **registration of the datasets with the IPO**. Attorney Cuyo suggested the use of the voluntary registration system to address various challenges related to artificial intelligence and copyright.

According to him, this system can serve as a tool for creators to assert their rights and for AI developers to verify the legitimacy of the datasets they use. It could also facilitate transparency and accountability as it requires the AI developers to ensure that their training datasets comply with copyright laws and that they respect the rights of the owners and rights holders. This proactive measure can likewise help mitigate disputes and foster a more ethical use of copyrighted materials in AI development.

Thus, while voluntary registration may not be mandatory, its strategic use can help navigate the complex interplay between copyright and artificial intelligence, ensuring that both creators' rights and the advancement of AI technologies are balanced and protected.

Fourth, the concept of **fair learning option** involves an independent entity that is responsible for evaluating training datasets and models used by AI systems. The entity would ensure that copyright owners are properly compensated and that their data is used in compliance with privacy and copyright laws. Attorney Cuyo highlighted that implementing a third-party oversight mechanism similar to a "*seal of good governance*" could provide transparency and accountability. This seal would act as a certification indicating that the AI developers have adhered to fair use standards and adequately compensated rights holders. Such a system would not only protect copyright holders but also build public trust in AI technologies by ensuring ethical and legal compliance.

Furthermore, this oversight could address data privacy concerns by verifying that personal data included in training datasets is handled responsibly and legally. By intertwining data privacy considerations with copyright protection, fair learning options would create a comprehensive framework for responsible AI development. This balance is crucial to fostering innovation while upholding the rights and privacy of individuals and creators.

Lastly, the implementation of digital receipts via blockchain. Relative to the previous two suggestions made by Atty. Cuyo is the

use of digital receipts via blockchain as also recommended by Atty. Angeles in ensuring that individual copyrights are protected in the use of copyrighted materials to AI models. Attorney Cuyo expressed:

We can start by requiring them na lahat ng materials nila na copyrighted there, there should be an indication [or] there should be a notice aside from the copyright notice, which [is] hindi mandatory, [na] pwedeng kailangan mandatory ilagay niya kung AI-generated. Pwedeng ganun siguro... That can be an immediate function.

According to Atty. Cuyo, this is a more straightforward initial step that would increase transparency and inform the public about the nature of the content they are engaging in. He further added that advancing from this, metadata or citations detailing the copyrighted materials used in the creation of AI-generated content could be included, enhancing traceability and accountability. This view is supported by Atty. Angeles who recommended the use of digital rights signature in the copyrighted materials as part of the author's or an owner's due diligence. He accentuated:

One thing that I can recommend, is that, maybe there can be a protocol where you embed a layer of digital right signature in the copyrighted material. So that when it is digitalized, it is caught by AI, parang AI will be adopted to catch the digital receipt. And then, parang maki-carry over sya on the process of AI.

This is very conceptual because... it's theoretically possible but you will have to agree on a universal system of digital information. Parang it is the same way as DocuSign enables or any e-signature service allowed to create a blockchain signature na ito yung provenance ng signature that is genuine, na nanggaling talaga ito kaya Atty. Angeles, kasi ito yung account niya, ito yung computer niya, at this point in time [he] gave his consent. So, I think that can be a practical application of blockchain. So, you embed a blockchain-based digital signature on the copyrighted work. Maybe the copyright office can be the one to give the digital signature din.

According to Atty. Angeles, the digital signature, potentially blockchain-based, would act as a verifiable digital receipt. When AI processes these materials, the digital signature would be carried

forward, maintaining a record of the work's provenance. Such a system would require a universal protocol, akin to how e-signature services operate, ensuring that each digital signature is authenticated and traceable.

Additionally, Atty. Angeles discussed that at the international level, the World Intellectual Property Organization (WIPO) could facilitate this system by establishing guidelines for digital signatures, creating a standardized identification mechanism. Individual states, therefore, could develop their own digital signature systems, adhering to WIPO's framework. This approach would ensure global consistency in identifying and verifying the provenance of copyrighted works.

AI systems would then be able to detect these digital signatures, integrating them into their output and providing a transparent account of the sources used. This method not only safeguards the rights of copyright holders but also promotes ethical AI development by ensuring that all inputs are properly attributed and compensated. This is a conceptual but feasible solution which underscores the importance of due diligence in protecting creative works in the digital age.

Ultimately, the suggestions and recommendations provided by the participants revolve around the need to ensure that the three important aspects of copyright are protected – *credit, compensation, and consent*.

Firstly, creators must be acknowledged for their work, which is why it is vital for AI developers to disclose the materials used. This transparency ensures proper credit is given. *Secondly*, creators should be compensated if their work is utilized. Determining the value and compensation, however, is challenging due to the complexity of algorithms, which makes it difficult to quantify the extent of usage of a specific copyrighted work in the AI's output.

Lastly, and equally important, is the aspect of consent. Creators should have the right to decide whether they want their work included in AI training datasets. They must be able to opt-out if they do not wish

for their materials to be used. This respect for the creator's choice can be facilitated through a voluntary copyright registration system. Such a system would allow creators to clearly state their preferences, thereby giving them control over the inclusion of their work in AI applications. This approach could be a significant step towards addressing the concerns of authors and creators in the evolving AI landscape.

CHAPTER VI

SUMMARY, CONCLUSION, AND RECOMMENDATIONS

This chapter presents the summary of the findings derived from the study concerning the liabilities of the AI actors in the unauthorized uses of copyrighted materials in AI models. It likewise outlines the researcher's recommendations for shaping future legislation geared toward addressing the complex issues concerning AI and copyright law.

Balancing Copyright and AI: Navigating Liability in the Digital Age

The rapid advancements in artificial intelligence have introduced significant copyright concerns, particularly issues concerning the liabilities of AI actors in the unauthorized uses of copyrighted materials as data sets in training AI models. Since AI systems rely heavily on extensive data sets for training and validation through machine learning techniques, examination of the interplay between AI technologies and copyright laws has become indispensable for developing legal frameworks that can adapt to the evolving nature of creative expression and knowledge production.

In this study, the researcher delved into four areas—global AI legislation, possible exemptions in the copyright law, the responsibility of AI actors in the unauthorized uses of copyrighted materials in AI models, and improvements in the Intellectual Property Code of the Philippines in the age of AI.

The examination international AI frameworks and legislations revealed that each country has a distinct approach to addressing copyright concerns and the liabilities of AI actors in the unauthorized uses of copyrighted materials. Some countries have specific provisions, while others rely on existing copyright laws to handle AI-related issues. AI legislation often permits the use of copyrighted materials for training data, under certain conditions, but does not extensively address the liabilities of AI actors. Nonetheless, the

following were identified as responsibilities and liabilities of AI developers arising from the permissible acts of using copyrighted materials as training data: promoting transparency and data disclosure, compliance with the copyright laws, ensuring data quality and integrity, and maintaining accountability measures. These measures help mitigate potential copyright infringements and foster a legally compliant and ethical AI ecosystem.

The exploration of AI actors' liability in copyright matters likewise highlights dual responsibilities during both the training (input) and utilization (output) stages of AI applications. The study reveals that AI developers and users have clearly defined liabilities based on their actions at each stage of the AI process, supported by participant narratives and case law references. It concludes that primary responsibility for unauthorized use of copyrighted materials lies with the developers, including data collectors and model trainers, while users may also face direct liability for infringing acts. This emphasizes the need for developers to ensure compliance with copyright laws from data collection onwards to mitigate legal risks. Additionally, the fair use provision serves as a crucial defense mechanism, illustrating the complex nature of copyright issues in AI.

Various permissible acts outlined in international AI legislation serve as exemptions to the general copyright law requirement of securing permission or authority before using copyrighted materials in a particular work. These permissible acts include exemptions for transparency requirements for GPAI models under open-source licenses, text and data mining, use of data for non-professional and scientific research, use of data for non-enjoyment, use of data for computer data processing, incidental use of data for efficient operation, and training data for creating new knowledge. These exemptions ensure that the use of copyrighted materials for AI training does not infringe on rights or diminish the value of the original works. This approach likewise aligns with the views of the participants who expressed support in AI training with copyrighted materials under specific conditions to respect copyright laws. These provisions foster technological progress while safeguarding creator's

interests, providing a comprehensive framework for AI development that respects intellectual property rights.

Exceptions to the permissible acts were also identified in the study. These encompass the option to reserve rights, ensuring fair competition, attribution to copyright holders, avoiding unreasonable harm, incidental and minimal use, and considering purpose and interference.

While international law provisions allow the use of copyrighted materials as training data, it is crucial to safeguard copyright holders' rights by identifying and enforcing limitations to these permissible acts. The challenge, however, lies in effectively enforcing these rights and ensuring compliance which requires robust mechanisms and regular audits. Hence, the participants of the study highlighted a few recommended measures to ensure that these protections are effectively enforced. These measures include valuation of copyright holders' work, disclosure requirements, registration of datasets with IPO, fair learning option, and digital receipts via blockchain.

In essence, the recommendations offered by the participants center on guaranteeing the protection of three fundamental aspects of copyright – credit, compensation, and consent.

Copy, Right? Proposals for Harmonizing Copyright and AI

In the dynamic landscape of technological innovation, the convergence of artificial intelligence and copyright law presents a myriad of challenges and opportunities. As AI systems become increasingly proficient in generating, analyzing, and manipulating digital content, the researcher proposes the following recommendations.

First, integration of provisions specific to machine learning into the Intellectual Property Code of the Philippines (IPC). This includes clearly defining permissible acts, exceptions, and responsibilities of AI developers. Prime considerations should focus on obtaining consent,

ensuring fair compensation, and giving proper credit to original creators, owners, and rights holders.

In line with this, the researcher introduced amendments to the IPC, as amended by Republic Act No. 10372, specifically addressing the limitations of copyright and the fair use provision of the copyrighted works. The amendments also include the insertion of a new chapter in Part IV of the said Code to comprehensively regulate the use of copyrighted materials in the context of artificial intelligence and machine learning. The full text of these proposed amendments is annexed at the end of this study.

Second, creation of the Implementing Rules and Regulations (IRR) to further empower the Intellectual Property Office (IPO) in exercising its administrative and adjudicatory powers to manage the repository of the copyrighted materials. The IRR should outline the practical measures for effectively enforcing the individual rights provided by the IPC to owners, creators, and rights holders of the copyrighted materials. Practical measures may include requirements for disclosing data, guidelines for registration of datasets with the IPO, and promoting fair learning practices. These measures should be geared toward ensuring proper attribution and compensation to where it is due, and securing consent from the rightful owners and copyright holders.

Third, embedding digital rights signature layer in copyrighted materials to trace provenance and facilitate proper attribution of the materials. This would enable the AI systems to detect and carry over these digital signatures during the content creation process. Implementing such a system would require consensus on universal protocol for digital information, akin to how other services use blockchain to verify the authenticity and provenance of signatures. This approach could be facilitated at the World Intellectual Property Organization (WIPO) level through a supplemental convention on copyright for digital provenance, creating international identification system. Thereafter, once WIPO established a convention, individual states could develop their own digital signature mechanisms within

this framework. The digital signature should include details such as the name and a blockchain-based verification protocol to ensure clear identification. AI systems should be capable of recognizing these signatures, integrating them into their outputs to provide transparency and traceability in the creation process, effectively protecting intellectual property rights in the digital age.

ANNEX

Proposed Recommendations to the Amendment of Republic Act No. 8293

Republic of the Philippines
HOUSE OF REPRESENTATIVES
Quezon City

NINETEENTH CONGRESS
_____ Regular Session

HOUSE BILL NO. _____

AN ACT AMENDING REPUBLIC ACT NO. 8293, OTHERWISE KNOWN AS THE "INTELLECTUAL PROPERTY CODE OF THE PHILIPPINES," TO INCLUDE A NEW CHAPTER ON THE USE OF COPYRIGHTED WORKS FOR MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE DEVELOPMENT

EXPLANATORY NOTE

The rapid advancement of Artificial Intelligence (AI) and Machine Learning (ML) has brought transformative changes across various sectors, including healthcare, education, industry, and creative arts. These technologies heavily rely on large datasets, which may include copyrighted materials, to learn and function effectively. However, the current framework of the Intellectual Property Code of the Philippines does not adequately address the use of copyrighted works in the context of AI and ML development.

This bill proposes the inclusion of a new chapter in the Intellectual Property Code to address the lawful use of copyrighted materials in AI systems. It is informed by comparative analysis of international AI

legislation, expert interviews, and academic research. The bill identifies specific permissible uses—such as text and data mining, scientific research, and incidental uses—while setting clear limitations and accountability measures to protect the rights of copyright holders.

In alignment with the principle of liberty and prosperity, this legislative measure promotes responsible innovation by ensuring that AI development does not come at the expense of the moral and economic rights of Filipino creators. It introduces obligations for AI developers and users, enhances transparency, and empowers the Intellectual Property Office of the Philippines (IPOPHL) to provide oversight and implement appropriate regulations.

This amendment aims to position the Philippines as a forward-looking, innovation-friendly jurisdiction that respects intellectual property rights while enabling technological advancement. The enactment of this legislation is imperative to ensure that our legal system remains responsive, equitable, and supportive of both innovation and creativity in the digital age.

Republic of the Philippines
HOUSE OF REPRESENTATIVES
Quezon City

NINETEENTH CONGRESS
First Regular Session

HOUSE BILL NO. _____

AN ACT AMENDING REPUBLIC ACT NO. 8293, OTHERWISE KNOWN AS THE "INTELLECTUAL PROPERTY CODE OF THE PHILIPPINES," TO INCLUDE A NEW CHAPTER ON THE USE OF COPYRIGHTED WORKS FOR MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE DEVELOPMENT

Be it enacted by the Senate and House of Representatives of the Philippines in Congress assembled:

SECTION 1. Section 184 of Republic Act No. 9283, as amended by Republic Act No. 10372, is hereby further amended to read as follows:

Section 184. *Limitations on Copyright.* - 184.1. Notwithstanding the provisions of Chapter V, the following acts shall not constitute infringement of copyright:

x x x

(c) The use of copyrighted works for training machine learning models is permitted under the following conditions:

(1) The use is non-commercial and for research or academic purposes.

(2) The data is lawfully acquired and appropriately attributed.

(3) The work is transformed or anonymized to a degree that prevents market substitution.

(d) Commercial AI developers must obtain licenses to use copyrighted works in training datasets, unless such use falls under existing fair use provisions.

x x x

SECTION 2. Section 185 of Republic Act No. 9283, as amended by Republic Act No. 10372, is hereby further amended to read as follows:

Section 185. *Fair Use of a Copyrighted Work.* - 185.1. The fair use of a copyrighted work for criticism, comment, news reporting, teaching including multiple copies for classroom use, scholarship, research, and similar purposes is not an infringement of copyright. Decompilation, which is understood here to be the reproduction of the code and translation of the forms of the computer program to achieve the inter-operability of an independently created computer program with other programs may also constitute fair use.

x x x

185.3 The following are exceptions to the permissible AI use, and shall not be allowed unless specifically licensed:

(a) Use of entire copyrighted works in datasets without transformation or anonymization.

(b) Use of unpublished works.

- (c) Uses that result in a substantial market harm to the original work.
- (d) Generation of outputs that closely mimic protected works (e.g., deepfake replicas, style transfer that infringes on distinctive artist styles).

SECTION 3. A new Chapter is hereby inserted in Part IV of Republic Act No. 8293, as amended, to be designated as Chapter IX, to read as follows:

**CHAPTER IX
MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE**

Section 191. *Declaration of Policy.* – The State recognizes the value of fostering innovation through emerging technologies such as Artificial Intelligence (AI) and Machine Learning (ML), while ensuring the protection and promotion of the rights of authors and creators. In line with the constitutional principles of liberty and prosperity, this Chapter aims to balance technological advancement with the integrity of original creative works, thereby supporting inclusive economic growth and responsible innovation.

Section 192. *Definitions.* – For purposes of this Chapter:

- (a) *Artificial Intelligence* (AI) refers to systems that simulate human intelligence processes including learning, reasoning, and self-correction.
- (b) *Machine Learning* (ML) refers to a method of data analysis that automates analytical model building through algorithms that learn from data.

(c) *Training Data* refers to digital content, including possibly copyrighted materials, used in the training or optimization of AI or ML systems.

(d) *Responsible AI Use* means the ethical, transparent, and lawful application of AI consistent with the rights of creators and the public interest.

Section 193. *Permissible Uses of Copyrighted Works for Machine Learning.* – Subject to the limitations set forth under this Chapter, the following acts may be deemed permissible:

- (a) Text and Data Mining (TDM) for analysis and research;
- (b) Use of copyrighted works for non-professional and scientific research;
- (c) Use of works for educational and non-personal enjoyment;
- (d) Use in computer data processing for system optimization;
- (e) Incidental or minimal use for operational efficiency;
- (f) Use for the creation of new knowledge through computational means;
- (g) Model training for purposes aligned with national development goals.

Section 194. *Limitations.* – The permissible uses of copyrighted works for Artificial Intelligence (AI) training under Section 193 shall be subject to the following conditions:

- (a) Authors shall retain the right to reserve or withhold the use of their works from inclusion in AI training datasets, thereby allowing them to exercise full control over whether or not their intellectual property is used for such purposes.
- (b) The use of copyrighted materials in AI training must not result in unfair competition or cause distortions in the

market that would undermine the rightful economic interests of original creators or copyright holders.

- (c) Where technically and practically feasible, proper attribution must be provided to the copyright holders of the materials used in AI systems, thereby ensuring recognition of creative contributions.
- (d) The use of copyrighted works must avoid causing unreasonable prejudice to the legitimate moral or economic interests of the rights holder, and should not compromise the integrity or value of the original work.
- (e) The scope of use must be limited to what is incidental and necessary for the effective development or function of the AI system, and shall not extend to uses that are excessive or unrelated to the stated objective.
- (f) The use of any copyrighted work must remain compatible with the declared purpose and function for which the AI system is designed, ensuring alignment with ethical and lawful use standards.

Section 195. *Responsibilities of AI Developers and Users.* —

(a) AI developers, data collectors, and model trainers shall:

- (1) ensure transparency by fully disclosing whether copyrighted materials are used in their systems, as well as identifying the specific sources of such materials.
- (2) maintain high standards for data quality and integrity throughout the development and operation of AI models, ensuring that the data used is accurate, reliable, and lawfully obtained.

- (3) implement accountability mechanisms, including the maintenance of audit trails and compliance monitoring systems, to ensure traceability and adherence to intellectual property regulations.
 - (4) required to uphold legal and ethical standards at every stage of the AI lifecycle – from the ingestion of training data to the generation of outputs – ensuring that no rights are violated in the process.
 - (5) promote transparency in the curation of datasets, with special emphasis on identifying and managing the inclusion of copyrighted content.
 - (6) establish and maintain a mechanism for disclosing the identity of works used in training datasets and for ensuring that proper credit is given to the original creators.
 - (7) must facilitate fair and equitable compensation mechanisms for creators whose copyrighted works are used in AI training or output generation, where such compensation is warranted.
 - (8) respect the autonomy and preferences of creators by providing them with a meaningful opportunity to opt in or opt out of having their works included in AI training datasets.
- (b) Primary liability for unauthorized or infringing use during the training phase shall rest with developers, data collectors, and model trainers.
 - (c) Users of AI-generated outputs may be held directly liable for infringing acts where intent or negligence is established.

Section 196. *Rights to Credit, Compensation, and Consent.* – To safeguard the rights of creators in the age of AI:

- (a) Creators shall have the right to be credited for the use of their works in AI systems, and mechanisms shall be developed to ensure attribution;
- (b) Fair compensation models shall be established to reward creators whose works contribute to AI training, recognizing the difficulty of quantifying contributions;
- (c) A voluntary copyright registration system shall be created by the IPOPHL, allowing creators to express preferences regarding the inclusion of their works in AI training datasets, thereby enabling meaningful consent.

Section 197. *Application of Fair Use Doctrine.* – The doctrine of fair use under Section 185 of this Act shall apply to AI and ML in accordance with the following factors:

- (a) The purpose and character of the use, including its transformative nature;
- (b) The nature of the copyrighted work;
- (c) The amount and substantiality of the portion used;
- (d) The effect on the market or potential market for the original work.

Section 198. *Transparency and Oversight.* – The Intellectual Property Office of the Philippines (IPOPHL) shall:

- (a) Issue guidelines on responsible AI practices;
- (b) Maintain a registry of publicly disclosed datasets used in AI development;

- (c) Facilitate stakeholder engagement and multi-sectoral oversight;
- (d) Promote open licensing and collective rights management mechanisms;
- (e) Develop and administer the voluntary registration and consent preference system under Section 196(c);
- (f) Create Implementing Rules and Regulations (IRR) to operationalize the rights under this Chapter, including practical requirements such as dataset disclosure, registration of datasets, and fair learning practices;
- (g) Establish a digital rights signature framework for embedding metadata in copyrighted materials to trace provenance, facilitate attribution, and allow AI systems to detect and retain origin data during content generation.

Section 199. *Remedies and Enforcement.* – Copyright holders shall be entitled to remedies under this Act, including injunctive relief, damages, and administrative sanctions for unauthorized use of their works in AI systems.

Section 200. *Liberty and Prosperity Clause.* – The provisions of this Chapter shall be construed in accordance with the fundamental principle that the law must advance both liberty and prosperity. The promotion of technological progress shall not diminish the rights, dignity, and rewards due to authors and creators.

SECTION 4. *Renumbering and Conforming Amendments.* – The subsequent Chapters and Sections of Republic Act No. 8293 are hereby renumbered accordingly. All references to previous Section numbers in existing provisions shall be read as referring to the renumbered Sections.

SECTION 5. *Implementing Rules and Regulations.* – The IPOPHL shall, within six (6) months from the effectivity of this Act, promulgate the necessary rules and regulations to implement the provisions of this Chapter.

SECTION 6. *Separability Clause.* – If any provision of this Act is declared unconstitutional or invalid, the other provisions not affected thereby shall remain in full force and effect.

SECTION 7. *Repealing Clause.* – All laws, presidential decrees, executive orders, and rules and regulations inconsistent with this Act are hereby repealed or modified accordingly.

SECTION 8. *Effectivity.* – This Act shall take effect fifteen (15) days after its publication in the Official Gazette or in a newspaper of general circulation.

APPROVED.

BIBLIOGRAPHY

PRIMARY AUTHORITIES

A. Treaties

Berne Convention for the Protection of Literary and Artistic Works, September 9, 1886, revised at Paris July 24, 1971, 1161 U.N.T.S. 3.

B. Other International Legal Instruments

AI Verify Foundation and Infocomm Media Development Authority [IMDA]. *Propose Model AI Governance Framework for Generative AI: Fostering a Trusted Ecosystem*, January 16, 2024. https://aiverifyfoundation.sg/downloads/Proposed_MGF_Gen_AI_2024.pdf. (last accessed May 30, 2024).

Australian Government: Department of Industry, Science and Resources. *Supporting responsible AI: discussion paper*, January 17, 2024. <https://consult.industry.gov.au/supporting-responsible-ai>. (last accessed May 30, 2024).

Commonwealth of Australia. *Safe and Responsible AI in Australia Consultation: Australian Government's interim response*. Australian Government: Department of Industry, Science, and Resources. Commonwealth of Australia, 2024. https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/safe-and-responsible-ai-in-australia-governments-interim-response.pdf. (last accessed May 30, 2024).

CSET. *Artificial Intelligence Law of the People's Republic of China (Draft for Suggestions From Scholars)* Center for Security and Emerging Technology. Center for Security and Emerging Technology, May 2,

2024. <https://cset.georgetown.edu/publication/china-ai-law-draft/>

Department for Science, Innovation, and Technology [DSIT]. *A pro-innovation approach to AI regulation: government response: Presented to Parliament by the Secretary of State for Science, Innovation and Technology by Command of His Majesty on 6 February 2024.*" <https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response#summary-of-consultation-evidence-and-government-response>. February 6, 2024. (last accessed May 29, 2024).

European Commission, Artificial Intelligence Act: European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (2019-2024) (EU), https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

Hiroshima AI Process Comprehensive Policy Framework, May 2, 2024. <https://www.soumu.go.jp/hiroshimaaiprocess/en/index.html#:~:text=After%20continuous%20discussions%20including%20at,code%20of%20conduct%20aimed%20at.>

Ministry of Internal Affairs and Communications Ministry of Economy, Trade and Industry. AI Guidelines for Business Ver 1.0 (April 19, 2024) (JP). https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240419_9.pdf

Secretary of State for Digital, Culture, Media and Sport by Command of Her Majesty. (2021). *National AI Strategy*. In HM Government (ISBN 978-1-5286-2894-5). Office for Artificial

Intelligence.

https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf (last accessed June 2, 2024)

Matt Sheehan and Carnegie Endowment for International Peace [Carnegie]. *China's AI Regulations and How They Get Made*. Carnegie Endowment for International Peace Publications Department, July 10, 2023. https://carnegie-production-assets.s3.amazonaws.com/static/files/202307-Sheehan_Chinese%20AI%20gov-1.pdf. (last accessed May 29, 2024).

World Intellectual Property Office (WIPO). *Conversation on Intellectual Property (IP) and Artificial Intelligence (AI): Revised Issues Paper on Intellectual Property Policy and Artificial Intelligence- Second Session*, WIPO/IP/AI/2/GE/20/, December 13, 2019.

C. Foreign Statutes

A bill to establish the National Artificial Intelligence Initiative, and for other purposes. House of Representatives 6216, (Mar. 12, 2020) (US), <https://www.congress.gov/116/bills/hr6216/BILLS-116hr6216ih.pdf>

Copyright Act of Japan, Act No. 48 of 1970, as amended by Act No. 30 of 2018.

Directive (EU) 2019/790 of the European Parliament and of the Council on Copyright and Related Rights in the Digital Single Market and Amending Directives 96/9/EC and 2001/29/EC of 17 April 2019 (July 5, 2019) [hereinafter Directive (Eu) 2019/790].

U.S. Copyright Office, *Compendium of U.S. Copyright Office Practices* § 107 (3d ed. 2021).

D. Domestic Law

An Act Defining Cybercrime, Providing for the Prevention, Investigation, Suppression and the Imposition of Penalties Therefor and for other purposes, Republic Act No. 10175, September 12, 2012.

An Act Prescribing the Intellectual Property Code and Establishing the Intellectual Property Office, providing for its Powers and Functions and for other purposes, Republic Act no. 8293 (June 6, 1997), as amended by Republic Act No. 10372, March 22, 2013.

An Act Protecting Individual Personal Information in Information and Communications Systems in the Government and the Private Sector, creating for this purpose a National Privacy Commission, and for other Purposes, Republic Act 10173, August 15, 2012.

An Act Providing for The Recognition and Use of Electronic Commercial and Non-Commercial Transactions, Penalties for Unlawful Use Thereof, and other purposes, Republic Act No. 8792, June 14, 2000.

E. Jurisprudence

a. United States

Authors Guild v. Google, Inc., No. 13-4829 (2d Cir. 2015)

Designers Guild v Russell Williams [2001] 1 All ER 700.

Hawkes v Paramount Film services [1934] Ch 593.

hiQ Labs, Inc. v. LinkedIn Corp., 938 F.3d 985 (9th Cir. 2019)

HRH Prince of Wales v Associated Newspapers [2008] IP & T 583.

Roberto Mata v. Avianca, Inc., 22-cv-1461 (PKC), U.S. District Court for the Southern District of New York, June 22, 2023.

b. Philippines

ABS-CBN Corporation v. Felipe Gozon, et. al., G.R. No. 195956, March 11, 2015

Filipinas Broadcasting Network, Inc. v. Ago Medical and Educational Center-Bicol Christian College of Medicine, G.R. No. 141994, January 17, 2005.

Filipino Society of Composers, Authors and Publishers, Inc. (FILSCAP) v. Anrey, Inc., G.R. No. 233918, August 09, 2022.

Pacita I. Habana, Alicia L. Cinco and Jovita N. Fernando V. Felicidad C. Robles and Goodwill Trading Co., Inc., G.R. No. 131522, July 19, 1999.

Pearl and Dean (Philippines), Inc. v. Shoemart, Inc., G.R. No. 148222, August 15, 2003.

Reyes v. AGB Nielsen Media Research Philippines, Inc., G.R. No. 176579, October 17, 2008.

F. Foreign Deliberation

Marcus Evans, Wilson Ang, Jeremy Lua, and Jeremiah Chew. *Singapore proposes Governance Framework for Generative AI*. Data Protection Report. February 4, 2024. <https://www.dataprotectionreport.com/2024/02/singapore-proposes-governance-framework-for-generative-ai/>.

SECONDARY AUTHORITIES

A. Books

Enrico Bonadio and Luke McDonagh. *Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the*

Consequences of Algorithmic Creativity. ERPN: Intellectual Property (2020).

Christophe Geiger. *Copyright as an Access Right, Securing Cultural Participation Through the Protection of Creators' Interests*. (2015): 73-109. <https://doi.org/10.22459/WIWCRC.01.2017.03>

John Creswell and David Creswell. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 5th ed. SAGE Publications, 2017.

I. Polishchuk and V. Koloniuk. *To The Question Of Defining The Concept Of Forging Art Works*. (2020): 672-684. <https://doi.org/10.33994/kndise.2020.65.66>.

Terry Hutchinson. (2010). *Researching and writing in law* (3rd ed.). Thomson Reuters Lawbooks, Australia.

B. Journal Articles

A. Borji. *A Categorical Archive of ChatGPT Failures*. ArXiv, abs/2302.03494 (2023). <https://doi.org/10.48550/arXiv.2302.03494>.

Andres Guadamuz. *The EU's Artificial Intelligence Act and Copyright* (July 01, 2024). *Journal of World Intellectual Property*, Forthcoming <http://dx.doi.org/10.2139/ssrn.4997854>

Benjamin Sobel. *A Taxonomy of Training Data: Disentangling the Mismatched Rights, Remedies, and Rationales for Restricting Machine Learning*. *Electronic* (2020). <https://doi.org/10.2139/ssrn.3677548>

Bohan Shi and Nazura Abdul Manap. *Research On Copyright Problem of Artificial Intelligence Generations in China*. *Journal of Information System and Technology Management* (2023). <https://doi.org/10.35631/jistm.831003>.

Calvin N. Mooers. *Computer Software and Copyright*. ACM Comput. Surv., 7 (1975): 45-72. <https://doi.org/10.1145/356643.356647>

Disha Shrivastava, H. Larochelle, and Daniel Tarlow. *Repository-Level Prompt Generation for Large Language Models of Code*. (2022): 31693-31715. <https://doi.org/10.48550/arXiv.2206.12839>.

E. Vavilin. *The status of artificial intelligence: from object to the subject of legal relations*. Vestnik Tomskogo Gosudarstvennogo Universiteta. Pravo (2022). <https://doi.org/10.17223/22253513/45/10>.

Eleonora Rosati. *Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law*. European Journal of Risk Regulation (2024). <https://doi.org/10.1017/err.2024.72>.

G. Spindler. *Copyright Law and Artificial Intelligence*. IIC - International Review of Intellectual Property and Competition Law, 50 (2019): 1049 - 1051. <https://doi.org/10.1007/s40319-019-00879-w>

George E. Newman and P. Bloom. *Art and authenticity: the importance of originals in judgments of value*. Journal of experimental psychology. General, 141 3 (2012): 558-69. <https://doi.org/10.1037/a0026035>.

J. A. T. Silva. *Fair Use in Post-publication Peer Review*. Journal of Educational and Social Research, 5 (2015): 13. <https://doi.org/10.5901/JESR.2015.V5N3P13>.

Jae-Hyun Kim. *A Study on the Copyright issues related to AI (Artificial Intelligence) creation in the era of the 4th industrial revolution*. Journal of Next-generation Convergence Information Services Technology (2021). <https://doi.org/10.29056/jncist.2021.08.06>.

Jessica Gillotte. *Copyright Infringement in AI-Generated Artworks* (June 2020). UC Davis Law Review, Vol. 53, No. 5, 2020, Available at SSRN: <https://ssrn.com/abstract=3657423>

Katherine Lee, Feder Cooper, & James Grimmelmann. *Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain* (July 27,

2023). Forthcoming, *Journal of the Copyright Society* 2024.
<http://dx.doi.org/10.2139/ssrn.4523551>

L. Ray Patterson and Stanley W. Lindberg. *The Nature of Copyright: A Law of Users' Rights*. The Michigan Law Review Association (1991).
<https://doi.org/10.2307/1289437>

Mark A. Lemley and B. Casey. *Fair Learning*. Intellectual Property: Other eJournal (2020). <https://doi.org/10.2139/ssrn.3528447>

Martin Kretschmer, Thomas Margoni and Pinar Oruc. *Copyright Law and the Lifecycle of Machine Learning Models*. IIC - International Review of Intellectual Property and Competition Law, 55 (2024): 110-138. <https://doi.org/10.2139/ssrn.4670563>.

Matthew J. Sag. *Orphan Works as Grist for the Data Mill*. Law & Literature eJournal (2012). <https://doi.org/10.2139/SSRN.2038889>

Milagros del Corral. *Copyright in the Digital Environment*. International Information & Library Review, 29 (1997): 201-204.
<https://doi.org/10.1006/IILR.1997.0048>

Pedro Martins Fernandes. *AI Training and Copyright: Should Intellectual Property Law Allow Machines to Learn?* *Bioethica* (2024). <https://doi.org/10.12681/bioeth.39041>.

Peter Henderson, Xuechen Li, Dan Jurafsky, Tatsunori Hashimoto, Mark A. Lemley and Percy Liang. *Foundation Models and Fair Use*. ArXiv, [abs/2303.15715](https://arxiv.org/abs/2303.15715) (2023).
<https://doi.org/10.48550/arXiv.2303.15715>.

Rita Matulionyte. *Australian Copyright Law Impedes the Development of Artificial Intelligence: What Are the Options?* IIC - International Review of Intellectual Property and Competition Law, 52 (2020): 417 - 443. <https://doi.org/10.1007/s40319-021-01039-9>.

Royani, Royani, Sondang Deri Maulina, Sugiyono Sugiyono, Rio Wahyudin Anugrah, and Brigitta Callula. 2024. *Recent Developments in Healthcare Through Machine Learning and Artificial Intelligence*. IAIC Transactions on Sustainable Digital Innovation (ITSDI) 6 (1):86-94. <https://doi.org/10.34306/itsdi.v6i1.680>.

Safet Emruli, Agim Nuhiu and B. Kadriu. *Copyright and Copyright Protection*. *European Journal of Interdisciplinary Studies*, 6 (2016): 36-40.

S. McIntyre. *Private Rights and Public Wrongs: Fair Use as a Remedy for Private Censorship*. *Entrepreneurship & Law eJournal* (2012). <https://doi.org/10.2139/SSRN.1998513>.

Virendra Ahuja. *Artificial intelligence and copyright: Issues and challenges*. *ILI Law Review*, Winter Issue 2020, <https://ssrn.com/abstract=3864922>

Y. Bysaga, D. Byelov and V. Zaborovskyi. *Artificial intelligence and copyright and related rights*. *Uzhhorod National University Herald*. Series: Law (2023). <https://doi.org/10.24144/2307-3322.2022.76.2.47>.

Zheng Yuan, Hongyi Yuan, Chuanqi Tan, Wei Wang, Songfang Huang and Feiran Huang. *RRHF: Rank Responses to Align Language Models with Human Feedback without tears*. *ArXiv*, abs/2304.05302 (2023). <https://doi.org/10.48550/arXiv.2304.05302>.

C. Other Publications

Aviv Hertzal Gaon. *Artificially Intelligent Copyright: Rethinking Copyright Boundaries*. Dissertation, York University, 2019.

Patrizia Anne Garcia. "Liberty in Creativity and Innovation, Prosperity in Protection Under the Rule of Law." *Thesis*, De La Salle Lipa, 2022.

Rachel Lois Gella. "Defining the Legal Role and Liability of Artificial Intelligence in Digital Consent towards Liberty and Prosperity in the Philippines." *Thesis*, University of St. La Salle, 2024.

Rustam Mutallimzada. *Liability for Copyright Infringement: An Investigation of the Legal Use of Trained Artificial Neural Networks in the Context of Copyright Law*. Master Thesis, Lund University, 2020.

D. Internet Sources

Artemio V. Panganiban, *Unleashing Entrepreneurial Ingenuity*, 12th General Assembly of the ASEAN Law Association, Makati Shangri-la Hotel, Makati City, February 26, 2015, available at <https://cjpanganiban.com/2015/02/26/unleashing-entrepreneurial-ingenuity/> [last accessed 11 May 2025].

Australian Government: Department of Industry, Science and Resources. *Supporting responsible AI: discussion paper*, January 17, 2024. <https://consult.industry.gov.au/supporting-responsible-ai>. (last accessed May 30, 2024).

Copyright. *NBER International Seminar on Macroeconomics* 8, no. 1 (2011): iv–iv. <https://doi.org/10.1086/667618>.

Hemant Arya and Tarun Kumar Bhatt. *Introduction of intellectual property rights*. (2021). <https://10.1016/B978-0-12-821471-8.00022-2>

Teresa Scassa, 'How Clearview AI Could Violate Copyright Law: Clearview AI scraped countless copyright-protected images from social media sites to develop a commercial facial recognition technology. Could copyright law be used to dismantle its service?', (*CIGI*, 10 March 2020), <https://www.cigionline.org/articles/how-clearview-ai-could-violate-copyright-law> (last accessed 10 December 2023).

Thomas Margoni. *Text and Data Mining in Intellectual Property Law: Towards an Autonomous Classification of Computational Legal Methods*

(May 16, 2020). CREATE working paper 01/2020 - Forthcoming in Calboli I. & Montagnani L., Handbook on Intellectual Property Research, Available at SSRN: <https://ssrn.com/abstract=3602699>