AI v copyright

Dermawan, Artha

Published in:
Journal of Intellectual Property Law and Practice

DOI:
10.1093/jiplp/jpad111

Published: 19.01.2024

Document Version
Publisher's PDF, also known as Version of record

Citation for pulished version (APA):
Dermawan, A. (2024). AI v copyright: how could public interest theory shift the discourse? *Journal of Intellectual Property Law and Practice*, *19*(1), 55-63. https://doi.org/10.1093/jiplp/jpad111

AI *v* copyright: how could public interest theory shift the discourse?

Artha Dermawan*

1. Introduction

Thanks to rapid progress in generative artificial intelligence (GenAI), the world has entered an era when technology, copyright and philosophy intersect in interesting ways. The advent of GenAI has subjected copyright law around the world to scepticism and opposition on many grounds, and a philosophical approach to the issue could stimulate a meaningful discussion on the aims, objectives and scope of copyright, ultimately leading to potential reforms and adaptations in the digital age. Sitting squarely at the centre of this intersection are large language models (LLMs), with the introduction of Bidirectional Encoder Representations from Transformers (BERT), Generative Pre-trained Transformer (ChatGPT), DALL-E 3, MusicGen and others.[1] In general, LLMs can be

The author

- Artha Dermawan is an Indonesia-qualified lawyer; a doctoral student at the University of Lapland, Finland; funded by the Max Planck Institute for Innovation and Competition, Germany; and a visiting researcher at the Intellectual Property Research Institute of Australia, the University of Melbourne. His research interests include the role of intellectual property as an engine for sustainable innovation in artificial intelligence, the pharmaceutical industry, the digital music industry and the traditional knowledge of indigenous and local communities.

Abstract

- To achieve a balance between rightsholders and the public in the age of generative AI (GenAI), it is necessary to redefine the concept of public interest in copyright law. This equilibrium is critical because it guarantees that rightsholders are rewarded for their work while also allowing for the dissemination and access of knowledge along with cultural expression. Additionally, redefining the concept of public interest in copyright could address emerging issues such as fair use, open access and the democratization of information in the digital age.
- The article examines how public interest in copyright would look in the development of GenAI by using Virginia Held's typology of public interest theory as the sole compass to address the question. The article finds that public interest in copyright could be adjusted to protect rightsholders, sustain cultural production and simultaneously accommodate the development of GenAI by first balancing the preponderance of individual interest or being supported by the preponderance of numbers and empirical terms; second, by making it consistent with individuals' overall interests and being agreed upon by the polity; and third, by judging the public interest in copyright on 'valid judgment' and normative content, where the judgment is based.
- While different typologies may be used to conceptualize public interest, the article suggests that it is important to explicitly identify what copyright law seeks to achieve and to balance the rights of creators with the public's interest in accessing creative works. This could be achieved through evidence-based analysis, agreement among individuals without conflicts and normative content based on the morality of society.

* Email: Artha.dermawan@ip.mpg.de. Open Access funding enabled and organized by the Max Planck Institute for Innovation and Competition. The author's sincere appreciation goes to the anonymous reviewers and the editorial team of the Journal of Intellectual Property Law and Practice. The author also would like to thank the following: Rosa Ballardini, Péter Mezei and Daria Kim for their helpful suggestions and comments on the previous draft. The responsibility for the content of this article and any errors herein remains with the author.

1 BERT is 'a method of pre-training language representations. Pre-training refers to how BERT is first trained on a large source of text, such as

doi:https://doi.org/10.1093/jiplp/jpad111
Advance Access Publication 29 December 2023

defined as: 'generative mathematical models of the statistical distribution of tokens in the vast public corpus of human-generated text, where the tokens in question include words, parts of words, or individual characters including punctuation marks'.[2] LLMs are generative because users can get samples from them by inquiring a specific fragment of text and ChatGPT shall provide what words are likely to come next with the help of Natural Language Processing called 'Reinforcement Learning from Human Feedback'.[3]

The discourse surrounding GenAI and copyright is a complex and multifaceted one. Can one train GenAI with copyright-protected works? Should we protect the outputs generated by it? and how much human creative contribution is necessary for the outputs generated by GenAI to be copyrightable?[4] are the questions that have been intensely debated around the world. On the question concerning the training data (input), legal discourses focus largely on the interpretation of fair use and text and data mining (TDM) exceptions.[5] Often the discussion reach the conclusion that the use of copyrighted

Wikipedia'. Google Cloud, 'Getting Started with the Built-in BERT Algorithm'. Available at https://cloud.google.com/ai-platform/training/docs/algorithms/bert-start (accessed 10 November 2023). ChatGPT is 'trained to follow an instruction in a prompt and provide a detailed response'. OpenAI, 'Introducing ChatGPT'. Available at https://openai.com/blog/chatgpt (accessed 10 November 2023). DALL-E 3 is 'an AI system that can create realistic images and art from a description in natural language'. Available at https://openai.com/research/dall-e (accessed 10 November 2023). MusicGen is 'a single Language Model (LM) that operates over several streams of compressed discrete music representation, ie, tokens'. Jade Copet, *et al,* 'Simple and Controllable Music Generation' (arXiv, 2023). Available at https://arxiv.org/abs/2306.05284 (accessed 10 November 2023).

2 Murray Shanahan 'Talking About Large Language Models' (Arxiv.org, 2023) 2. Available at https://arxiv.org/pdf/2212.03551.pdf (accessed 10 November 2023). See, Philipp Hacker, et al, 'Regulating ChatGPT and other Large Generative AI Models' (arXiv, 2023). Available at https://arxiv.org/pdf/2302.02337.pdf (accessed 10 November 2023).

3 Reinforcement Learning from Human Feedback is 'an area in machine learning research that incorporates human guidance or feedback to learn an optimal policy'. Zihao Li, Zhuoran Yang and Mengdi Wang, 'Reinforcement Learning with Human Feedback: Learning Dynamic Choices via Pessimism' (arXiv, 2023). Available at https://arxiv.org/pdf/2305.18438.pdf (accessed 10 November 2023). See, further research on similar topics, including (but not limited to) Nisan Stiennon, *et al*, 'Learning to summarize with human feedback'. (Advances in Neural Information Processing Systems 33, 2020); Long Ouyang, *et al*, 'Training Language Models to Follow Instructions with Human Feedback' (arXiv, 2022). Available at https://arxiv.org/abs/2203.02155 (accessed 10 November 2023). Paul F Christiano, *et al*, 'Deep Reinforcement Learning from Human Preferences'. Available at https://proceedings.neurips.cc/paper_files/paper/2017/file/d5e2c0adad503c91f91df240d0cd4e49-Paper.pdf (accessed 10 November 2023).

4 A number of legal scholars have expressed their views on the issue, such as: Pamuela Samuelson, 'Allocating Ownership Rights in Computer-Generated Works' (47 *University of Pittsburgh Law Review* 1185–8, 1986); Annemarie Bridy, 'Coding Creativity: Copyright and the Artificially Intelligent Author (5 *Stanford Technology Law Review* 1–28, 2012); Emmanuel Salami, 'AI-generated Works and Copyright Law: Towards a Union of Strange Bedfellows' (vol. 16(2) *Journal of Intellectual Property Law & Practice* 124–35, 2020); James Grimmelmann, 'There's No Such Thing as a Computer-Authored Work – And It's a Good Thing, Too' (39 *Columbia Journal of Law & the Arts* 403–16, 2016); Robert Denicola, 'Ex Machina: Copyright Protection for Computer-Generated Works' (69 *Rutgers University Law Review* 251–86, 2016); Amir H. Khoury, 'Intellectual Property Rights for Hubots: On the Legal Implications of Human-like Robots as Innovators and Creators' (35 *Cardozo Arts & Entertainment Law Journal* 635–68, 2017); Robert Yu, 'The Machine Author: What Level of Copyright Protection Is Appropriate for Fully Independent Computer-Generated Works?' (165 *University of Pennsylvania Law Review* 1245–50, 2017); Enrico Bonadio, et al, 'Intellectual Property Aspects of Robotics" (9 *European Journal of Risk Regulation* 655–60, 2018); Daryl Lim, 'AI & IP: Innovation & Creativity in an Age of Accelerated Change' (52 *Akron Law Review* 814–72, 2018); Enrico Bonadio and Nicola Lucchi (eds.), *Non-Conventional Copyright: Do New and Non Traditional Works Deserve Protection?* (Cheltenham, Edward Elgar, UK, 1–520, 2018); Mark Lemley and Bryan Casey, 'Remedies for Robots' (86 *University of Chicago Law Review* 1311–96, 2019); Peter Yu, 'The Algorithmic Divide and Equality in the Age of Artificial Intelligence' (72 *Florida Law Review* 331–88, 2020); Dan L. Burk, 'Thirty-Six Views of Copyright Authorship, by Jackson Pollock' (58 *Houston Law Review* 263, 2020); Enrico Bonadio and Luke McDonagh, 'Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity' (2 *Intellectual Property Quarterly* 112, 2020); Tim W. Dornis, 'Artificial Creativity: Emergent Works and the Void in Current Copyright Doctrine' (22 *Yale Journal of Law & Technology* 1, 2020); Vincenzo Iaia, 'To Be, or Not to Be … Original Under Copyright Law, That Is (One of) the Main Questions Concerning AI-Produced Works' (Vol. 71(9), *GRUR International*, 793–812, 2022); Péter Mezei, 'You AIn't Seen Nothing yet' – Arguments against the Protectability of AI-generated Outputs by Copyright Law' in Maurizio Borghi and Roger Brownsword (eds.), *Informational Rights and Informational Wrongs: A Tapestry for Our Times* (Routledge, Abingdon, 126–43, 2023).

5 TDM generally refers to the process of obtaining valuable information from massive amounts of data. See, including (but not limited to) Artha Dermawan,' Text and Data Mining Exceptions in the Development of Generative AI Models: What the EU Member States Could Learn From the Japanese "nonenjoyment" Purposes?' (Forthcoming in *The Journal of World Intellectual Property*, 1–18, 2024); Benjamin Sobel, 'A Taxonomy of Training Data Disentangling the Mismatched Rights, Remedies, and Rationales for Restricting Machine Learning.' In Jyh-An Lee, Reto Hilty and Kung-Chung Liu (eds.), Artificial Intelligence & Intellectual Property (Oxford, Oxford University Press, UK, 222–3, 2021); Christophe Geiger,' The Missing Goal-Scorers in the Artificial Intelligence Team: of Big Data, the Fundamental Right to Research and the Failed Text and Data Mining Limitations in the CSDM Directive' (*PIJIP/TLS Research Paper Series* No. 66, 383–94, 2021); Christophe Geiger and Vincenzo Iaia,' The Forgotten Creator: Towards a Statutory Remuneration Right for Machine Learning of Generative AI' (Forthcoming in *Computer Law & Security Review,* 1–15, 2023); Eleonora Rosati, 'No Step-Free Copyright Exceptions: The Role of the Three-step in Defining Permitted Uses of Protected Content (including TDM for AI-Training Purposes)' (Stockholm Faculty of Law Research Paper Series No. 123, 1–23, 2023); Rossana Ducato and Alain Strowel, 'Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to "Machine Legibility' (50 *International Review of Intellectual Property and Competition Law*, 649–84, 2019); Rossana Ducato and Alain Strowel,' Ensuring Text and Data Mining: Remaining Issues with the EU Copyright Exceptions and Possible Ways Out' (43(5) *European Intellectual Property Review*, 322–7, 2021); Sean Flynn, et al, 'Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action' (42(7) *European Intellectual Property Review* 393–8, 2020); Josef Drexl, et al,'Artificial Intelligence and Intellectual Property Law Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate' (*Max Planck Institute for Innovation and Competition Research Paper* No. 21–10, 1–26, 2021); Tatsuhiro Ueono,' The Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication' (70(2), *GRUR International*, 145–52, 2021); Thomas Margoni and Martin Kretschmer,' A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' (71(8), *GRUR International*, 685–701, 2022); Valentina Moscon, 'Data Access Rules, Copyright and Protection of Technological Protection Measures in the EU. A Wave of

materials in LLMs training should be first authorized by the rightsholders,[6] and the TDM exception should balance the interests of the rightsholders and the public.[7] The rationale behind the balance of interests is based on the broadly recognized notion that 'copyright contains free spaces to ensure follow-on creativity and to secure important fundamental rights and the public interest, in particular allowing research to be undertaken using protected material'.[8] That said, public interest is also increasingly used as a proxy for the interests of LLM researchers and users. There has been a built-in tension between the interests of the rightsholders and those of the LLM researchers and users.[9]

Giblin and Weatherall argue that 'there are obvious problems with conflating 'the public interest' with one specific interest group or another. Such conceptions will rarely convince others in the context of the perennial copyright debate'.[10] In the case of GenAI, the unauthorised use of copyright-protected works is tantamount to 'taking authors out of the equation entirely takes the heart out of copyright law, and fail[ing] to recognise the public's interest in the creation of a diversity of cultural material'.[11] Policymakers might avoid these issues by recognizing that in the age of LLMs, the public interest must embrace a variety of purposes, and philosophical approaches might shine a light.

This article examines how public interest in copyright would look in the development of GenAI by using Held's typology of public interest theory as the sole compass to address the question. The article finds that public interest in copyright law could be adjusted to protect rightsholders, sustain cultural production and simultaneously accommodate the development of GenAI by first balancing the preponderance of individual interest or being supported by the preponderance of numbers and empirical terms; second, by making it consistent with individuals' overall interests and being agreed upon by the polity; and third, by judging the public interest in copyright on 'valid judgment' and normative content, where the judgment is based. While different typologies may be used to conceptualize public interest, the article contends that it is important to explicitly identify what copyright law seeks to achieve and to balance the rights of creators with the public's interest in accessing creative works. This could be realized through evidence-based analysis, agreement among individuals without conflicts and normative content based on the morality of the society. The article further argues that these three approaches to defining public interest in copyright are not mutually exclusive and could be used in combination.[12]

2. Public interest and copyright: searching for common ground

The relationship between copyright and public interest is as close as two coats of paint. The legitimacy of copyright law as a longstanding guardian of the public interest dates back to the introduction of the Statute of Anne in 1710.[13] The Statute of Anne, the first copyright law of modern history, has inspired many legislative traditions around the world such as the Canadian Copyright Act which is designed as 'a balance between promoting the public interest in the encouragement and dissemination of works of the arts and intellect and obtaining a just reward for the creator (…)'.[14]

Propertisation of Information' (Max Planck Institute for Innovation and Competition Research Paper No. 23–14, 1–24, 2023).

6 See, Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, OJ L 130, 17.5.2019 where Art 3 of the Directive provides a TDM exception for the acts of 'reproductions and extractions made by research organisations and cultural heritage institutions to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access'. Moreover, Art 4 of the Directive provides an exception for reproductions and extractions of lawfully accessible works/subject matter for TDM to provide significant legal certainty for both private and public entities undertaking TDM. See also, the European Union (EU) AI Act Proposal where LLMs will be required to disclose the training data being used. The press release is Available at https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai (accessed 11 November 2023).

7 Charlotte Gerrish and Anders M. Skavlan, 'European Copyright Law and Text and Data Mining Exceptions and Limitations: In Light of the Recent DSM Directive, is the EU Approach a Hindrance or Facilitator to Innovation in the Region?' (2 *Stockholm Intellectual Property Law Review* 58, 2019) *See*, Carys J. Craig, 'The AI-Copyright Challenge: Tech-Neutrality, Authorship, and the Public Interest' (2021) in Ryan Abbott (ed.) *Research Handbook on Intellectual Property and Artificial Intelligence* (Cheltenham, Edward Elgar, UK, 134–55, 2022).

8 Sean Flynn, et al, n.5, 393–8. See, Ruth L. Okediji (ed), *Copyright Law in an Age of Exceptions and Limitations* (New York, Cambridge University Press, NY, USA, 2017). From the European Union context, see eg Christophe Geiger and Franciska Schönherr, 'Defining the Scope of Protection of Copyright in the EU: The Need to Reconsider the Acquis Regarding Limitations and Exceptions' in Tatiana E. Synodinou (ed.), *Codification of European Copyright Law: Challenges and Perspectives* (Zuid-Holland, Kluwer Law International, the Netherlands, 133, 2012).

9 This is of concern for the free and unrestricted access by LLMs to copyright works that are used for the advancement of LLMs.

10 Rebecca Giblin and Kimberlee Weatherall, 'If We Redesigned Copyright From Scratch, What Might it Look Like?' in Rebecca Giblin and Kimberlee Weatherall (eds), *What if We Could Reimagine Copyright?* (Australian national University Press, 2017) 6. Available at https://press-files.anu.edu.au/downloads/press/n2190/pdf/book.pdf (accessed 10 November 2023).

11 *Ibid.*

12 For example, a copyright law that is both empirically effective and aligned with the general interests of society may also be seen as morally justifiable.

13 The Statute of Anne (1710) was designed for 'an act for the encouragement of learning, by vesting the copies of printed books in the authors of purchaser of such copies (….)' Available at https://case.edu/affil/sce/authorship/statueofanne.pdf (accessed 10 November 2023).

14 LTC Harms, *The Enforcement of Intellectual Property Rights: A Case Book* (3rd edn WIPO 2012).

Gervais provides that public interest has always been part of international copyright law and policy.[15] The statement makes perfect sense given that international copyright treaties have also recognized the public interest as a guiding principle. For instance, Article 7 of the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement)[16] adopts the public interest objectives of IP protection and the principle ensures that IP protection should 'promote the public interest in sectors of vital importance to their socio-economic and technological development'.[17] Okediji argues that the negotiators and drafters of the TRIPS Agreement were particularly mindful of the public interest principles on which copyright is based.[18] Moreover, a reference to IP protection and public interest can be seen in the preamble of the WIPO Copyright Treaty,[19] which recognizes the need to maintain a balance between the protection of IP and the promotion of creativity and innovation. The treaty also acknowledges the importance of limitations and exceptions to copyright for educational, research and other public interest purposes.

Moreover, in copyright law, the rights of those who are protected, namely authors and owners, are frequently used interchangeably with the public interest, and the Berne Convention appears to have treated the terms 'public interest' and 'public good' in this manner.[20] In connection with the development of LLMs, there is ongoing debate about whether copyright law should prioritize the interests of rightsholders or the public. The necessity of redefining the concept of public interest in copyright is required to establish a balance between these competing interests. As copyright law was initially created to protect the rightsholders as GenAI progresses, there is a likelihood that copyright could be used to foster innovation and overlook the exclusive rights of the creators, and on the other hand, it could also be used to restrict access to information and discourage innovation. Therefore, it is important to consider the broader concept of the public interest in copyright law, as Feintuck also suggests that a key factor in determining the public interest is ensuring that policies and regulations promote the well-being of society as a whole, rather than just benefiting a select few.[21]

However, the question remains: how does one define public interest in copyright law? The concept may have existed long before the Statute of Anne was introduced but philosophers, academics and lawyers have not been able to define the concept and often left undefined.[22] Sorauf included public interest in the list of most ambiguous phrases and complained that scholars could not reach consensus on 'what they are trying to define: a goal, a process, or a myth'.[23] Schubert argues that the concept of public interest 'makes no operational sense, notwithstanding the efforts of a generation of capable scholars'.[24] If there is no exact definition of public interest, why not simply ignore the concept? One must realize that the concept of public interest is central to the democratic systems, as outlined by Giblin and Weatherall:

> The idea that 'we' are not just a welter of self-interested individuals out to further our own specific interests regardless of the impact on others, but rather, that we make up a society, in which guise we can have a shared set of interests. It also suggests that, on occasion, furtherance of those shared interests can and should take precedence over individual or group self-interest.[25]

15 Daniel Gervais, 'Fair Use, Fair Dealing, Fair Principles: Efforts to Conceptualize Exceptions and Limitations to Copyright' (57 *Journal of the Copyright Society of the USA,* 499–520, 2010).

16 Agreement on Trade-Related Agreements of Intellectual Property Rights, art 13, 15 April 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 33 I.L.M. 1197 (1994). Article 7 of the TRIPS Agreement states the public interest objectives as follows:

> The protection and enforcement of intellectual property rights should contribute to the promotion of technological innovation and to the transfer and dissemination of technology, to the mutual advantage of producers and users of technological knowledge and in a manner conducive to social and economic welfare, and to a balance of rights and obligations.

See, Peter K. Yu, 'The Objectives and Principles of the TRIPS Agreement' (46 Hous. L. Rv. 979, 1007, 2009) argues that Art 7 is 'intended to strike a balance that more widely promotes social and economic welfare'. See also, Haochen Sun, 'Copyright Law as an Engine of Public Interest Protection' (2019) 16 *Northwestern Journal of Technology and Intellectual Property* 123. Available at https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1318&context=njtip (accessed 10 November 2023).

17 Article 8(1) of the TRIPS Agreement.

18 Ruth L. Okediji, 'Toward an International Fair Use Doctrine' (39 *Columbia Journal of Transnational Law 75*, 81–82, 2000).

19 World Intellectual Property Organization Copyright Treaty, 20 December 1996, 36 I.L.M. 65 (1997).

20 Rebecca Giblin and Kimberlee Weatherall, n 10, 5.

21 Mike Feintuck, *The Public Interest in Regulation* (Oxford, Oxford University Press, UK, 2004).

22 Rebecca Giblin and Kimberlee Weatherall, *ibid*, 3. See, Robert A. Dahl and Charles E. Lindblom, *Politics, Economics, and Welfare: Planning and Politico-Economic Systems Resolved into Basic Social Processes* (New YorkHarper & Row, NY, USA, 1963) 501.

23 Frank Sorauf, 'The Conceptual Muddle' in Carl J Friedrich (ed), *Nomos V: The Public Interest* (California, Atherton Press, USA, 1962) 190. See, Rebecca Giblin and Kimberlee Weatherall, *ibid,* 4. See also, Isabella Alexander, *Copyright Law and the Public Interest in the Nineteenth Century* (Oxford and Portland, Oregon: Hart Publishing, 2010) 16, noting that 'the notion of "public interest" is not a single or unified concept – its content will vary depending upon who is considered to make up "the public" and who is articulating its interests. At times different interests may come into conflict, and at other times they may be complementary'.

24 Rebecca Giblin and Kimberlee Weatherall, *ibid*, 4 citing Glendon Schubert, *The Public Interest: A Critique of the Theory of a Political Concept* (Westport, Connecticut, Praeger Publishers Inc, USA,1982) 224.

25 Rebecca Giblin and Kimberlee Weatherall, *ibid*, 8 citing Charles J. Fox and Hugh T. Miller, *Postmodern Public Administration: Toward Discourse* (Thousand Oaks, California: Sage Publications, USA, 1995) 123–124.

To redefine the concept of public interest in copyright law and to make it fit with technological advancement, one needs to give the concept some content, and luckily, the tools of philosophy could help.[26] Philosophical theories such as Held's public interest conception on legal and political philosophy (Held's typology) could provide guidance in determining what constitutes public interest in the context of copyright law and could further the definition of public interest.

3. Held's typology of public interest, copyright and LLMs: synchronizing the triangle

There are several public interest typologies which has been previously presented by scholars including Banfield,[27] Schubert,[28] Cochran[29] and; Leys and Perry.[30] However, this article will focus on Held's typology because it recognizes that individuals have a variety of interests, including personal, moral, social and political interests.[31] Moreover, Held's typology emphasizes the importance of balancing competing interests, and might be relevant on the issue of GenAI as to promote the common good which is suitable for copyright law. Held's typology is a framework that categorizes different types of public interests conceptions into three categories: (i) Preponderance Theories, (ii) Common Interest Theories and (iii) Unitary Theories.[32] The following subsections will discuss each typology and its significance to copyright and LLMs or GenAI in general.

3.1 Preponderance theories

The preponderance theory suggests that the interests of the general public should be given greater weight than the interests of individuals or corporations.[33] The preponderance way is often applied in legal cases where there is a conflict between individual rights and the greater good of society. Emblazoned by the works of Hobbes and Bentham, the theories suggest that public interest 'cannot be in conflict with a preponderance or sum of individual interests'.[34]

This preponderance seems to have been indirectly applied to copyright, given that similar concepts can be found in the copyright exceptions and limitations or fair use doctrine.[35] As Davies in *Copyright and Public Interest* argues:[36]

> Whether a particular act is 'in the public interest' is probably not subject to any objective tests. Inherent in the noble motive of the public good is the notion that, in certain circumstances, the needs of the majority override those of the individual, and that the citizen should relinquish any thoughts of self-interest in favour of the common good of society as a whole.

Unfortunately, in light of current technological advancements and the rise of LLMs, the adherents of this theory need to answer the following: How should this preponderance be evaluated? and how should minority interests be taken into account? As Giblin and Weatherall have questioned: 'is it by simple majority of numbers, or weighted by political strength? or should it be: "judged in empirical or behavioral terms, as a higher degree of force, or a greater weight of actual opinion, or a superior group strength?"'[37] To answer the first question, it seems that, for the time being, the most plausible way to judge is through the preponderance of numbers, empirical terms and consensus. To answer whether the use of copyright-protected works in training GenAI systems or by granting prompt-based copyright is part of the majority or minority interests in this game would require empirical terms and the actual opinion of the artist and the art community in general. Empirical research on the magnitude of LLMs' influence on the rightsholders and the economy in general is also one of the important things to establish.

3.2 Common interest theories

As the name suggests, the hallmark of this theory is unanimity.[38] This means that all members of a group must agree on a decision before it can be made. The common interest theories suggest 'that something will only be in the public interest where it is in the interests of all members of a polity'.[39] This approach can lead to more

26 Rebecca Giblin and Kimberlee Weatherall, *ibid*, 9.

27 Edward C. Banfield, 'Note on Conceptual Scheme' in Martin Meyerson and Edward C. Banfield, *Politics, Planning and the Public Interest* (New York, NY: The Free Press, USA, 1955).

28 Glendon Schubert, n 24.

29 Cochran explores the relationship between the interest-based and common good–centred approaches from various perspectives. C E. Cochran 'Political science and the public interest' (36 *Journal of Politics* 327, 1974).

30 Wayne AR Leys and Charner M. Perry, 'Philosophy and the Public Interest: A document' (Paper presented at the Symposium of the Western Division of the American Philosophical Association, University of Wisconsin, 1 May 1959).

31 Virginia Held, *The Public Interest and Individual Interests* (New York, NY: Basic Books, USA, 1970).

32 *Ibid.* See, Rebecca Giblin and Kimberlee Weatherall, *ibid*, 10.

33 Virginia Held, *ibid*, 19.

34 *Ibid*, 43.

35 Rebecca Giblin and Kimberlee Weatherall, *ibid*, 10.

36 Gillian Davies, *Copyright and the Public Interest* (2nd edn Sweet & Maxwell 2002) 7. Found in Rebecca Giblin and Kimberlee Weatherall, *ibid,* 10.

37 Giblin and Kimberlee Weatherall, *ibid*, 10.

38 Giblin and Kimberlee Weatherall, *ibid*, 11.

39 *Ibid*, 11. Citing Virginia Held, *ibid*, 44.

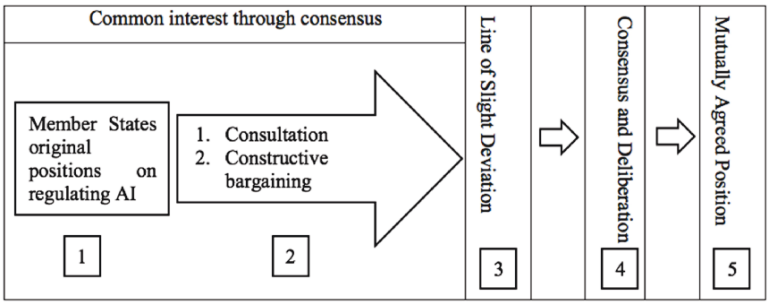

Figure 1. The process of consensus according to the ASEAN Way.[a]

[a]Based on a similar illustration in Artha Dermawan and Péter Mezei, 'Artificial Intelligence and Consensus-Based Remuneration Regime in Southeast Asia' (2023) 21. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4625850 (accessed 10 November 2023). See, Samuwel C. Padmakumara, 'A Conceptual Analysis on "ASEAN Way" as a Normative Approach for Conducting Regional Affairs' (2021) 9. Available at: https://fgs.cmb.ac.lk/wp-content/uploads/Vol-06-No-01-1-12.pdf (accessed 10 November 2023).

thorough and thoughtful decision-making processes, as all perspectives are taken into account. However, it can also be time-consuming and difficult to achieve in larger groups or when dealing with complex issues as the theories suggest that 'where a policy triggers conflict between individual interests it cannot be in the public interest'.[40] As illustrated by Fig. 1 below, one of the implementation in determining public interest through common interest is consensus which has long been practised by several Southeast Asian countries such as Indonesia through the concept of *gotong royong* or mutual assistance which has become a local wisdom and served as the foundation for political discourse on the nature of power and the features of life in Indonesia.[41] Furthermore, in the Association of Southeast Asian Nations (ASEAN), a similar concept is well-known as the ASEAN Way. As stipulated in the preamble and art. 2(2)(e) of the ASEAN Charter,[42] ASEAN adopted its own principle called the ASEAN Way, a fundamental principle of consensus during the decision-making.

In copyright law, unanimity may not be possible, as there may be differing opinions and interests among stakeholders such as creators, distributors and consumers. In such cases, compromises and the balancing of interests may be necessary to reach a consensus that is fair and reasonable for all parties involved. For example, when negotiating copyright laws for the use of copyright-protected works in training GenAI, rightsholders may want to ensure they receive fair compensation for their work, while LLMs researchers and providers may prioritize ease of access and affordability of training data for the models. In this case, a compromise may involve implementing a system where rightsholders are paid royalties based on the number of uses of their works in the respective models, while also ensuring that LLMs' researchers and providers have easy and affordable access to the copyright-protected works they want for the purpose of training the models. In this scenario, it seems that the theory is best applied if the public interest in copyright aims to protect human artists from the potential harm caused by GenAI.

Furthermore, in the above example, the public interest may not align with the interests of all members of a polity. For example, a policy that aims to remunerate rightsholders may be in the public interest, but some GenAI researchers and providers may resist it as it could make innovation even harder. As Barry argues, common interests of all members of the society rarely occur.[43] However, some examples of common interests might be possible in the future such as in the development of AI where humans desire for safety and security and this could be achieved through transparent access to AI training data.[44] As Noto La Diega *et al* argue, access to AI

40 Giblin and Kimberlee Weatherall, *ibid*, 11.

41 John R. Bowen, 'On the Political Construction of Tradition: Gotong Royong in Indonesia' (45 *The Journal of Asian Studies* 545–61, 1986).

42 The preamble of the ASEAN Charter reads 'RESPECTING the fundamental importance of amity and cooperation, and the principles of sovereignty, equality, territorial integrity, non-interference, consensus and unity in diversity'. Art 2(2)(e) of the Charter emphasizes that ASEAN and its member states shall act in accordance with the 'non-interference in the internal affairs of ASEAN member states'.

43 *Ibid*, 11. Citing B M. Barry, 'The Use and Abuse of the Public Interest' in Carl J Friedrich (ed), *Nomos V: The Public Interest* (California: Atherton Press, USA, 1962) 199.

44 See, the European Union (EU) AI Act Proposal where LLMs or AI providers will be required to disclose the training data being used for

training data could also be used for public good such as affordable access to medicine and sustainable patent governance.[45]

Based on the foregoing analysis, it seems that common interest theories are less suitable to be applied if copyright wishes to achieve a balance between protecting rightsholders and at the same time fostering innovation in the field of AI. According to Giblin and Weatherall, the preponderance and common interest theories are often portrayed as process theories because 'they provide a *process* for determining where the public interest lies without purporting to give it normative content. This has led to criticism for lack of usefulness'.[46] For example, in the context of determining whether using copyright-protected materials to train LLMs or GenAI models (as 'x') is in the interests of individuals, when referring to preponderance theories, we could consider Held's opinion as a guidance:[47]

> to assert that '*x* is in the interests of a preponderance of individuals' implies only that '*x* is in the interests of a preponderance of individuals.' … [I]f we want to know whether a given *x* is in the public interest, we want to know something else than the empirical fact that it is in the interests of a preponderance of individuals, although being in the interests of a preponderance of individuals may well be among the possible good reasons for believing that such an *x* is in the public interest.

To determine whether a particular copyright policy is in the interests of a majority or of the entire society, process theories must include some normative content.[48] For instance, when it is claimed that a certain policy is in the public interest because it is in the interests of all parties, a normative determination should be made. A process theory that lacks normative content may fail to account for the diverse values and interests of different groups within a society, leading to policies that are not truly in the public interest. Furthermore, a process theory that lacks normative content may also fail to address the issue of power imbalances between different groups in society.[49] Therefore, any process theory that aims to determine the public interest must take into account the complex and often competing values and interests of different groups, as well as the power dynamics that shape decision-making processes. Only then society can ensure that copyright policies truly serve the common good and promote a fair and equitable distribution of benefits. To sum up, it is important to recognize that copyright policies do not exist in a vacuum, but are shaped by broader social, economic and political contexts that must also be considered in any analysis of their impact on society.

3.3 Unitary theories: valid judgment and normative content

Plato and Aristotle both introduced unitary theories in the public interest for the first time, which propose that the well-being of the community should be prioritized over individual interests. Plato believed that a just society could only be achieved through a philosopher-king who would govern with wisdom and virtue, while Aristotle argued that a virtuous middle class should hold power.[50] These theories have influenced political philosophy for centuries and continue to shape discussions about the role of government in promoting the common good, including Hegel, Marx and Held.[51] Nevertheless, Held views that 'valid judgment' is one of the most critical benchmarks in determining public interest because 'if something is in the public interest as a matter of 'valid' judgment, it must also be in each individual's interest.

transparency reason. The proposal is available at https://www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2021/0206/COM_COM2021_0206_EN.pdf (accessed 10 November 2023).

45 This article reflects on what sustainable patent governance looks like in the age of the rise of AI and critically evaluates the relationship between AI, European patent law and sustainability with a focus on Sustainable Development Goal 9, ie, to build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation. In Section 2, the article focuses on AI datasets and suggests a recalibration that revolves around the concept of therapeutic data commons. Guido Noto La Diega, Gabriele Cifrodelli and Artha Dermawan, 'Sustainable Patent Governance of Artificial Intelligence: Recalibrating the European Patent System to Foster Innovation (SDG 9)' in Bita Amani, Caroline Ncube, and Matthew Rimmer (eds), *Elgar Companion on Intellectual Property and Sustainable Development Goals* (Cheltenham, Edward Elgar, UK, 2023).

46 Giblin and Kimberlee Weatherall, *ibid*, 11.

47 *Ibid*, 11. Citing Virginia Held, *ibid*, 84.

48 Giblin and Kimberlee Weatherall, *ibid*, 12.

49 For instance, a copyright policy that benefits large AI corporations but harms individual creators may be justified as being in the public interest, but in reality, it may simply reflect the interests of the powerful at the expense of the marginalized.

50 Aristotle describes how the relationship between the individual and the state should look like: 'For even if the end is the same for a single man and for a state, that of the state seems at all events something greater and more complete whether to attain or to preserve; though it is worthwhile to attain the end merely for one man, it is finer and more godlike to attain it for a nation or for city-states'. Aristotle. *Nicomachean Ethics* (The Internet Classics Archive). Available at http://classics.mit.edu/Aristotle/nicomachaen.html (accessed 10 November 2023). See, Plato, *Republic* (Penguin Books, 2003). See also, Kadri Simm, 'The Concepts of Common Good and Public Interest: From Plato to Biobanking' (2011) 20 Cambridge Quarterly of Healthcare Ethics 554.

51 Unitary theories in the public interest, such as those proposed by Hegel and Marx, emphasize the importance of a unified society that works towards the common good. These theories argue that individual interests should be subordinated to the needs of the community as a whole, and that social and economic inequalities should be eliminated to create a more just and equitable society. Julie Mostov, 'Karl Marx as Democratic Theorist' 22 (1989) Polity 195–212. *See*, Gary Browning, *A History of Modern Political Thought: The Question of Interpretation* (Oxford: Oxford University Press, UK, 2016) 25–45. See also, Aaron Berman, 'Haslanger, Marx, and the Social Ontology of Unitary Theory: Debating Capitalism's Relationship to Race and Gender' (8 *Journal of Social Ontology* 118–50, 2022).

Equally, if something is not in the interest of an individual as a matter of such judgment, then it cannot be in the public interest either'.[52]

Whether something fulfils the criteria of 'valid judgement' should be determined by the universal moral order, and whether something is in the public interest is determined on that basis. This means that actions which align with the moral order are considered valid or justifiable, while those that go against it are not. Additionally, the public interest is not determined solely by individual preferences or opinions, but rather by what is deemed morally right and beneficial for society as a whole. As Giblin and Weatherall have argued, how does one reach agreement on where this 'objective (morally) good' actually lies?[53] With regard to the use of copyright-protected works in training GenAI systems, one may wonder whether the use without the authorization of the copyright owner is morally acceptable. This question could be answered easily given that copyright infringement is an immoral act, at least from the society perspective.[54] On the other hand, in its application, unitary theories could also be problematic, given the impermissibility of conflict between public interest and individual interest.[55]

However, the concept of 'valid judgement' is not foreign to the copyright system. A prominent example comes from the Japanese 'non-enjoyment' purposes, which led to the introduction of the world's broadest TDM exceptions.[56] The purpose of 'enjoyment' and 'non-enjoyment' under Article 30–4 of the Japanese Copyright Law could also be a form of 'valid judgement', which culminates in whether an action fulfils moral requirements.[57] The act of 'enjoy' refers to the action of 'to accept and appreciate highly emotional things or physical interest, etc.'.[58] The purpose of 'enjoyment' under the Japanese Copyright Act is interpreted as 'enjoying the expression of a work by appreciating it through human senses'.[59] The determination of whether a person's interest, in the case of a rightsholder, is unduly compressed is determined by whether the activity undertaken aims to make others enjoy a work. Under the Japanese Copyright Law, one can observe that copyright protection of a work constitutes the public interest because the 'legitimate' assessment of 'enjoyment' should also be in the interest of every individual.

4. Conclusion: what does copyright seek to achieve?

This article has sought to contribute ideas to the copyright and AI debate by redefining the definition of public interest in copyright law through Held's typology. The aforementioned analysis provides a middle ground on how the public interest in copyright law might be moulded to suit the development of AI. This goal could be achieved through: (i) balancing it with the preponderance of individual interest and supported by the preponderance of numbers, consensus and empirical terms, which should include artists and the art community in order to determine the extent to which AI affects rightsholders and the copyright economy in general. A nuanced understanding of the potential benefits and drawbacks of these technologies is necessary to ensure that any copyright policies are grounded in evidence-based analysis; (ii) making it in line with the general interest of each individual and being agreed upon as a whole without any conflicts and (iii) the public interest in copyright could be based on 'valid judgement' and normative content where the judgement is based on the morality of the society. This

52 Rebecca Giblin and Kimberlee Weatherall, n 11, 12. Citing Virginia Held, *ibid*, 53.

53 *Ibid.*

54 There has been an increasing number of lawsuits against generative AI companies. To see lists of the various cases, see, 'Lawsuits v. AI.' Available at https://chatgptiseatingtheworld.com/2023/10/19/master-list-of-lawsuits-v-ai-chatgpt-openai-microsoft-meta-midjourney-other-ai-cos/ (accessed 10 November 2023). In response to these lawsuits, most GenAI companies, such as StabilityAI, Anthropic, Hugging Face, Adobe, Microsoft, Google, Apple and Meta, argued that they should not be required to compensate copyright holders for training GenAI models. See, Wes Davis, 'AI Companies Have All Kinds of Arguments Against Paying for Copyrighted Content.' (The Verge, 2023). Available at: https://www.theverge.com/2023/11/4/23946353/generative-ai-copyright-training-data-openai-microsoft-google-meta-stabilityai (accessed 10 November 2023).

55 Virginia Held, *ibid*, 156–158.

56 For detailed discussion on this topic, see Tatsuhiro Ueno, 'The Flexible Copyright Exception for "Non-Enjoyment" Purposes—Recent Amendment in Japan and Its Implication' (70(2) *GRUR International* 147, 2021). See also, Art 10 (1) of the Japanese Copyright Act.

57 The Article provides the following:

Art 30–4 (Exploitation without the Purpose of Enjoying the Thoughts or Sentiments Expressed in a Work)

It is permissible to exploit a work, in any way and to the extent considered necessary, in any of the following cases, or in any other case in which it is not a person's purpose to personally enjoy or cause another person to enjoy the thoughts or sentiments expressed in that work; provided, however, that this does not apply if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation:

(i) if it is done for use in testing to develop or put into practical use technology that is connected with the recording of sounds or visuals of a work or other such exploitation;

(ii) if it is done for use in data analysis (meaning the extraction, comparison, classification or other statistical analysis of the constituent language, sounds, images or other elemental data from a large number of works or a large volume of other such data; the same applies in Art 47–5, paragraph (1), item (ii));

(iii) if it is exploited in the course of computer data processing or otherwise exploited in a way that does not involve what is expressed in the work being perceived by the human senses (for works of computer programming, such exploitation excludes).

The Japanese Copyright Research and Information Center (CRIC), 'The Copyright of Japan (Official Translation)'. Available at https://www.cric.or.jp/english/clj/cl2.html (accessed 11 November 2023).

58 Tatsuhiro Ueno, *ibid*, 150.

59 *Ibid.*

article finds that in relation to the latter provision, it turns out that 'valid judgement' is not alien to the copyright system given that a similar concept already exists in Japan through the 'non-enjoyment' purposes.

Based on Held's typology and the analysis above, it can be suggested that it does not matter which typology to adopt in order to better conceptualize public interest in the advent of LLMs or AI. As Giblin and Weatherall contend, 'it is impossible to argue sensibly that any copyright policy is in the public interest without explicitly identifying what it seeks to achieve'.[60] So then, what exactly is copyright law seeking to achieve? One of the fundamental desires is to achieve a balance between protecting the rights of creators and promoting the public's interest in accessing creative works. It aims to encourage creativity and innovation by providing legal protection for original works while also allowing access to these works for educational, research and other purposes.[61]

However, to genuinely answer that question, particularly in the age of GenAI, one has to go back and revisit the copyright justifications. The copyright historical justifications can be roughly categorized into 'naturalist' and 'instrumentalist' theories.[62] These theories provide different perspectives on the purpose and scope of copyright law, with naturalist theories emphasizing the inherent rights of creators and instrumentalist theories focusing on the economic and social benefits of copyright protection.[63] Both approaches have influenced the development of copyright law over time and continue to shape debates around issues such as fair use, digital piracy and the balance between copyright and free speech. The inherent rights approach emphasizes the moral and ethical considerations of protecting creative works, while instrumentalist theories prioritize the practical benefits of copyright protection for society as a whole. Despite their differences, both approaches recognize the importance of balancing the interests of creators and users in order to promote innovation and creativity.[64]

In the development of AI, these debates are becoming even more complex as questions arise around the use of copyrighted works to train the AI and the ownership and protection of works created by machines. Whichever conception of copyright is used, at least everyone agrees that the objective is to balance the interests of creators, users, AI researchers and society as a whole. To summarize, the typology of public interest in copyright presented in this article will not only be useful in the discourse on the use of copyrighted works in AI research and development but can also be used by judges and policy makers in deciding all matters concerning AI, copyright (IP in general) and public interest in the near future.

60 Rebecca Giblin and Kimberlee Weatherall, n 11, 16.

61 Mark S. Nadel, 'How Current Copyright Law Discourages Creative Output: The Overlooked Impact of Marketing' (2004) 19 *Berkeley Technology Law Journal*, 785–856. See, Alain Strowel, 'Droit d'auteur et copyright, Divergences et Convergences' (Bruylant, 1993).

62 Rebecca Giblin and Kimberlee Weatherall, *ibid.*

63 *Ibid.*

64 Chelsea Bodimeade and Felicity Deane, 'Evolving theory of IP Rights: Promoting Human Rights in the Agreement on Trade-Related Aspects of Intellectual Property Rights' (18(8) *Journal of Intellectual Property Law & Practice* 603–14, 2023). See, Ruth G. Okediji, 'Copyright and Public Welfare in Global Perspective' (7 *Indiana Journal of Global Legal Studies* 117–89, 1999).