TANZANIAN COPYRIGHT LAW AND LARGE LANGUAGE MODELS: EXPLORING THE INTERSECTION

Albert Moshi, Tanzania Institute of Accountancy, P.O. Box 9522, Dar es Salaam, Tanzania

ABSTRACT

Large language models (LLMs) are at the forefront of artificial intelligence, transforming text creation in digital platforms. These models generate human-like text by leveraging patterns learned from large datasets. However, the emergence and rapid usage of large language models in various sectors presents significant challenges to Tanzania's copyright law[1], raising crucial questions about authorship, ownership, and the copyright infringement. The Act[2] provides the basis for the protection of copyright and related rights within the country, though it was enacted before the emergence of artificial intelligence technologies, including large language models (LLMs). This situation poses challenges on its applicability in the current digital ecosystem. This study aimed to analyse and identify gaps in the existing Tanzanian Copyright law and assess its capacity to address the implication introduced by large language models in text creation. The study revealed that existing Copyright law lacks the necessary provisions to effectively address the implications introduced by large language models. Key issues such as authorship, ownership, and copyright infringement emerged as critical points of concern. The study recommended a collaborative approach involving policymakers, legal experts, and technology developers to create the legal framework that can support the usage of large language models while protecting intellectual property rights.

---

[1] Copyright and Neighbouring Rights Act, 1999 (Act No. 7 of 1999).
[2] Copyright and Neighbouring Rights Act, 1999 (Act No. 7 of 1999).

1.0 Background

The capability of large language models (LLMs) to create human-like text has highlighted the critical importance of examining copyright laws. Large language models (LLMs) are advanced artificial intelligence systems that are capable of understanding user queries and generate text based on user prompts.[3] These models leverage machine learning algorithms and natural language processing (NLP) that extract and analyse vast volumes of data from diverse sources, allowing them to produce text content in response to user prompts.[4] Examples of large language models include ChatGPT (Generative Pre-trained Transformer) developed by OpenAI, which has several versions, including ChatGPT-3.0 and ChatGPT-4.0. Other models are BERT (Bidirectional Encoder Representations from Transformers) and T5 (Text-to-Text Transfer Transformer), both created by Google. Additionally, Claude AI model developed by Anthropic and POE AI (Platform for Open Exploration) developed by Quora. LLMs become powerful tools capable of generating text and performing a variety of language-related tasks, making them invaluable across various fields such as in business, agriculture, manufacturing, education and in entertainment.[5] For instance, in business and corporate environments, LLMs are used to summarize reports, making it easier for managers and supervisors to digest large amounts of information quickly. They also assist in preparing documents, such as proposals and presentations, thus saving time and reducing workload. In academic and research institutions, LLMs support publications by aiding in literature reviews, drafting articles, and even generating research summaries. Their ability to process and analyse vast amounts of text allows researchers to focus on other productive activities rather than wasting time in reading and writing reports. However, while LLMs offer significant benefits in streamlining text creation, they also create complex challenges for copyright laws worldwide.[6]

Globally, Copyright laws are designed to protect a wide range of creative works, ensuring that authors and creators have exclusive rights to their original expressions. As Large language models have advanced, copyright laws worldwide have created paradoxical situations regarding the implications brought by AI-generated content. For example, U.S. copyright law[7]

---

[3] Li, J., Tang, T., Zhao, W. X., Nie, J. Y., & Wen, J. R. (2024). Pre-trained language models for text generation: A survey. *ACM Computing Surveys*, *56*(9), 1-39.
[4] Wu, Y. (2024). Large Language Model and Text Generation. In *Natural Language Processing in Biomedicine: A Practical Guide* (pp. 265-297). Cham: Springer International Publishing.
[5] Johnsen, M. (2024). *Large Language Models (LLMs)*. Maria Johnsen.
[6] German, D. M. (2024). Copyright-Related Risks in the Creation and Use of ML/AI Systems. In *Large Language Models in Cybersecurity: Threats, Exposure and Mitigation* (pp. 145-152). Cham: Springer Nature Switzerland.
[7] Copyright Act of 1976 and Digital Millennium Copyright Act (DMCA) of 1998.

currently does not recognize AI as an author. United Kingdom, copyright law[8] provides protection to works created by natural person and does not explicitly address the protection of AI-generated contents. Similarly, South African copyright law[9] and India copyright law[10] does not directly address AI-generated texts which raises questions about authorship and ownership.

In Tanzania, Copyright law has historically focused on protecting traditional forms of creative expression, protecting the rights of authors and creators in various fields such as literature, music, and the arts as stipulated in section 3 and 5 of the Act.[11] However, the emergence of artificial intelligence (AI) presents significant challenges to this Act. AI technologies specifically LLMs are capable of generating text autonomously basing on user prompts, raising questions about authorship and ownership. This technological transformation in text creation raises concerns about the adequacy of traditional copyright protections which are rooted in human authorship. Without clear guidelines on the copyright protection of LLMs generated text, developers and users may face uncertainty regarding ownership and rights. This legal ambiguity can lead to disputes and a lack of confidence in the utilisation of LLMs in text creation. Thus, this study aimed to analyse and identify gaps in the existing Tanzanian Copyright law and assess its capacity to address the implication introduced by LLMs text generation.

2.0 Overview of AI powered Large Language Model and Copyright Protection

AI-powered large language models (LLMs) operate by processing user inputs to generate text outputs which can include project reports, research summaries, concept notes, work processes and guidelines and other various forms of text content.[12] The two key parties involved in text creation are the developers[13] and the users[14], each playing essential role in the creation and utilization of LLMs text content. **Developers** are responsible for designing and training LLM models, using advanced machine learning algorithms and vast datasets.[15] Conversely, **users**

---

[8] Copyright, Designs and Patents Act 1988
[9] Copyright Act 98 of 1978
[10] Copyright Act of 1957
[11] Copyright and Neighbouring Rights Act, No. 7 of 1999.
[12] Al-Shehri, K. (2023). The Synergy of AI and Large Language Models: Transforming Natural Language Processing. *Advances in Computer Sciences*, *6*(1).
[13] Organizations or individuals who create and train the Large Language models.
[14] Users may include businesses, educators, researchers, content creators, and the general public
[15] Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., ... & Mirjalili, S. (2023). A survey on large language models: Applications, challenges, limitations, and practical usage. *Authorea Preprints*.

interact with LLMs by providing prompts and receiving generated text content.[16] When an LLMs generates text, it raises complex questions regarding copyright ownership. Is the copyright belonging to the developers of the model, the users who provided the prompts, or does it exist in legal gray area? This ambiguity complicates the determination of ownership in the eyes of law and highlights the need for clarity in copyright law as it pertains to LLMs text content creation.

LLMs are trained on extensive and massive datasets curated by developers, which can include copyrighted material.[17] If the output generated by an LLMs closely mimics, reproduces, or resembles portions of copyrighted works, it could potentially constitute copyright infringement.[18] This raises the important question of who bears the risk: the users who interact with the model or the developers who created it, as both parties may inadvertently violate copyright laws through their engagement with LLMs text outputs. The machine learning algorithm processes that build up large language models are often complex making it difficult for users to trace how specific inputs have been processed to the desired text outputs.[19] This situation complicates the ability of users to responsibly declare the origins of the content, as they may not fully understand whether the output resembles existing works or if it is genuinely original. Consequently, this ambiguity raises significant ethical and legal questions about plagiarism.

3.0 Implication of AI powered large language models on Tanzania copyright law

3.1 Ownership of literary works

The emergence of Large Language Models (LLMs) presents a compelling dilemma regarding the ownership of literary works under Tanzania copyright law. LLMs challenges the Act as it blurs the lines of creative origin. Currently the Act offers protection across various forms of creative expression, including literary, artistic, and musical works as stipulated in section 3 and 5 of the Act. Its main purpose is to provide legal recognition and protection for original works, ensuring that creators have exclusive rights to reproduce, distribute, and adapt their content.

---

[16] Brade, S., Wang, B., Sousa, M., Oore, S., & Grossman, T. (2023, October). Promptify: Text-to-image generation through interactive prompt exploration with large language models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (pp. 1-14).

[17] Li, H., Deng, G., Liu, Y., Wang, K., Li, Y., Zhang, T., ... & Wang, H. (2024). Digger: Detecting copyright content mis-usage in large language model training. *arXiv preprint arXiv:2401.00676.*

[18] Thongmeensuk, S. (2024). Rethinking copyright exceptions in the era of generative AI: Balancing innovation and intellectual property protection. *The Journal of World Intellectual Property*, *27*(2), 278-295.

[19] Barbierato, E., & Gatti, A. (2024). The challenges of machine learning: A critical review. *Electronics*, *13*(2), 416.

Additionally, section 27 of the Act[20] addresses neighbouring rights, which protect the interests of performers, producers, and broadcasters. The emergence of AI-powered large language models (LLMs) presents significant implications on the Act[21], particularly as the existing legislation primarily focuses on human authorship and traditional forms of creative expression as section 3 of the Act[22] states *"Copyright shall subsist in original works of authorship, including literary, musical, and artistic works, provided that such works are expressed in a tangible form."* This implies that works must be original and expressed in a tangible medium to qualify for copyright protection, emphasizing the necessity of natural person authorship in the creation of such works. The author under section 2 of the Act[23] is defined as *"'Author' means the person who creates a work, including a literary, artistic or musical work."* By specifying that an author is a "person," the law implicitly excludes AI powered LLMs from being considered authors. This section clearly emphasizes that an author is a person who creates original works. This distinction is significant because it reinforces the idea that copyright protection is inherently tied to human creativity and expression. However, with the rise of LLMs capable of generating text content, the question of whether LLMs can also be considered an author has emerged. Currently, The Act[24] does not recognize LLMs as authors; they focus on human creators. As a result, any text content generated by LLMs would generally not have a designated author in the eyes of the law, leading to ambiguity regarding ownership. This gap raises important discussions about how copyright laws should evolve to address the reality of LLMs generated text content. When Large Language Model generates text, it raises complex questions regarding copyright ownership. Is the copyright attributed to the developers of the model, the users who provided the prompts, or does the resulting content inhabit a legal gray area? Without clear guidelines on the copyright status of LLMs generated text works, developers and users may face uncertainty regarding their rights, potentially leading to disputes over ownership and infringement.

3.2 Originality of literary works

Under the **Act**[25], the concept of originality is a fundamental criterion for copyright protection of literary works. Originality, in this aspect, means that a work must be the result of the author's

---

20 Copyright and Neighbouring Rights Act, No. 7 of 1999.
21 Copyright and Neighbouring Rights Act, No. 7 of 1999.
22 Copyright and Neighbouring Rights Act, No. 7 of 1999
23 Copyright and Neighbouring Rights Act, No. 7 of 1999
24 Copyright and Neighbouring Rights Act, No. 7 of 1999
25 Copyright and Neighbouring Rights Act, No. 7 of 1999

independent intellectual effort and must not be copied from existing works as stated in section 3 of the Act[26] as *"A work shall be considered original if it is the result of the author's own intellectual effort and is not a mere copy of another work."* The question of originality in LLMs generated works presents a compelling challenge.[27] It raises the question of whether the **originality of a work produced by LLMs belong to the developer of the AI model, or to the user who provides the prompts and context for text creation.** On one side, the developer of the LLMs has designed the algorithms and trained the model, thus contributing to its capability to generate text content. This raises the argument that they hold a degree of authorship over the outputs, as their technical expertise and choices shape the LLMs creative capability. On the other side, the user plays a crucial role by providing specific prompts and guiding the LLMs responses which significantly influence the nature and direction of the content produced, suggesting that users might claim a form of originality based on their input and intent. The Act[28] is in a state of turmoil, failing to clearly define the originality of works created at the intersection of human creativity and machine generated content. This challenge reflects broader questions about creativity and originality when using LLMs in generating content highlighting the urgent need for amending the Act[29], as traditional notions of originality must be re-evaluated in light of collaborative human-AI powered LLMs interactions.

Furthermore, the rise of large language models (LLMs) has exaggerated concerns about Plagiarism.[30] LLMs can generate text that closely resemble the existing works on which the model was trained, the risk of unintentional plagiarism increases, complicating the responsibilities of developers and users. Plagiarism as an offense is stipulated in Section 11 of the Act[31] which state *"Any person who, without the authorization of the owner of the copyright, does any of the following acts in relation to a work shall be deemed to have infringed the copyright in that work: (a) reproduces the work; (b) distributes copies of the work to the public; (c) performs the work in public; (d) communicates the work to the public."* This section emphasizes the importance of originality and the need to acknowledge sources, thereby protecting the rights of authors. However, When users interact with LLMs to generate text content, often lack awareness and knowledge of the algorithmic processes and datasets that

---

[26] Copyright and Neighbouring Rights Act, No. 7 of 1999
[27] Iaia, V. (2022). To be, or not to be… Original under copyright law, that is (one of) the main questions concerning AI-produced works. *GRUR International*, 71(9), 793-812.
[28] Copyright and Neighbouring Rights Act, No. 7 of 1999
[29] Copyright and Neighbouring Rights Act, No. 7 of 1999
[30] Plagiarism refers to the act of presenting another person's ideas, text, or creations as one's own, which can violate copyright laws if the original work is protected.
[31] Copyright and Neighbouring Rights Act, No. 7 of 1999

support the LLMs capabilities.[32] The LLMs may unintentionally produce outputs that closely resemble existing works, raising concerns about originality and copyright infringement.[33] Users often cannot trace how an LLMs arrived at a specific output, making it difficult to determine whether the text content is original or closely mimics existing copyrighted material. Users may not realize that what they perceive as unique and original may, in fact, be a reconfiguration of previously existing material. This can lead to unintentional plagiarism, where users might present LLMs generated content as their own without recognizing its derivative nature. The crucial question arises of who bears the legal risk? developers or users? From a legal perspective, developers create the algorithms and train the model on large datasets that enable LLMs to generate text content. They may bear responsibility for ensuring that their model do not produce infringing material. On the other side, users who interact with LLMs to generate text content also face potential risks as they may inadvertently produce text content that resemble existing copyrighted materials. The Act[34] under section 11 broadly defines copyright infringement without specifying whether liability falls solely on the user, the developer, or both. This means that anyone involved in the creation, distribution, or reproduction of a work that infringes on copyright could potentially face legal consequences. The absence of clarity regarding who bears the risk can lead to confusion and legal uncertainty. Users might assume that the LLMs, as a tool, frees them from copyright infringement, while developers may believe that their role is limited to providing the technology. This duality of risk highlights the need for clear legal frameworks that outline responsibilities and liabilities for both developers and users, ensuring that all parties understand their roles in preventing copyright infringement while fostering innovation.

4.0 Conclusion

Currently, Tanzania copyright law is fundamentally grounded in the concept of human authorship, which emphasises that only works created by natural person can be protected under copyright granting them exclusive rights to their expressions and preventing unauthorized use of their works. As a result, the copyright law face challenges in adapting to the realities of large language models generated content, which do not fit neatly into the traditional definitions of authorship and ownership. As LLMs continue to progress, it is vital to modify and revising

---

[32] Cheng, L., Varshney, K. R., & Liu, H. (2021). Socially responsible ai algorithms: Issues, purposes, and challenges. *Journal of Artificial Intelligence Research*, *71*, 1137-1181.
[33] Abbott, R., & Rothman, E. (2023). Disrupting creativity: Copyright law in the age of generative artificial intelligence. *Fla. L. Rev.*, *75*, 1141.
[34] Copyright and Neighbouring Rights Act, No. 7 of 1999

existing copyright law to address challenges brought by LLMs generated contents. This adaptation will be crucial for protecting intellectual property rights while fostering usage of large language models in content generation

4.2 Recommendations

To the policy makers

Policy makers are recommended to develop specific guidelines that clarify the rights and responsibilities of both developers and users of LLMs. These guidelines should aim to protect intellectual property while fostering innovation.

Policy makers should prioritize the reform of copyright law to address the unique challenges posed by LLMs generated content. This reform should focus on redefining concepts of authorship, originality, and ownership in light of LLMs. For instance, changing the legal definitions of authorship to include contributions from LLMs while also recognizing human input. Additionally, the clarification of ownership rights must be clearly defined, establishing who holds copyright over works generated by large language models (LLMs). This includes determining whether the rights belong to the user or the developer of the model.

Policy makers are also advised to develop a regulatory framework that specifically addresses the use of AI in creative industries. This framework should include guidelines on copyright protection for AI-generated works, ensuring that both creators and developers understand their rights and responsibilities.

To Law maker

Law makers are recommended to establish a mechanism for the regular review and updating of the Act to keep pace with technological advancements of LLMs and evolving practices in content creation, ensuring that the legislation remains relevant and effective.

Bibliography

Abbott, R., & Rothman, E. (2023). Disrupting creativity: Copyright law in the age of generative artificial intelligence. *Fla. L. Rev.*, *75*, 1141.

Al-Shehri, K. (2023). The Synergy of AI and Large Language Models: Transforming Natural Language Processing. *Advances in Computer Sciences*, *6*(1).

Barbierato, E., & Gatti, A. (2024). The challenges of machine learning: A critical review. *Electronics*, *13*(2), 416.

Brade, S., Wang, B., Sousa, M., Oore, S., & Grossman, T. (2023, October). Promptify: Text-to-image generation through interactive prompt exploration with large language models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (pp. 1-14).

Cheng, L., Varshney, K. R., & Liu, H. (2021). Socially responsible ai algorithms: Issues, purposes, and challenges. *Journal of Artificial Intelligence Research*, *71*, 1137-1181.

German, D. M. (2024). Copyright-Related Risks in the Creation and Use of ML/AI Systems. In *Large Language Models in Cybersecurity: Threats, Exposure and Mitigation* (pp. 145-152). Cham: Springer Nature Switzerland.

Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., ... & Mirjalili, S. (2023). A survey on large language models: Applications, challenges, limitations, and practical usage. *Authorea Preprints*.

Iaia, V. (2022). To be, or not to be… Original under copyright law, that is (one of) the main questions concerning AI-produced works. *GRUR International*, *71*(9), 793-812.

Johnsen, M. (2024). *Large Language Models (LLMs)*. Maria Johnsen.

Li, H., Deng, G., Liu, Y., Wang, K., Li, Y., Zhang, T., ... & Wang, H. (2024). Digger: Detecting copyright content mis-usage in large language model training. *arXiv preprint arXiv:2401.00676*.

Li, H., Deng, G., Liu, Y., Wang, K., Li, Y., Zhang, T., ... & Wang, H. (2024). Digger: Detecting copyright content mis-usage in large language model training. *arXiv preprint arXiv:2401.00676*

Li, J., Tang, T., Zhao, W. X., Nie, J. Y., & Wen, J. R. (2024). Pre-trained language models for

text generation: A survey. *ACM Computing Surveys*, *56*(9), 1-39.

Thongmeensuk, S. (2024). Rethinking copyright exceptions in the era of generative AI: Balancing innovation and intellectual property protection. *The Journal of World Intellectual Property*, *27*(2), 278-295.

Wu, Y. (2024). Large Language Model and Text Generation. In *Natural Language Processing in Biomedicine: A Practical Guide* (pp. 265-297). Cham: Springer International Publishing.