International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

# Intellectual Property Challenges in the Era of Big Data and Large Language Models

**Mohammad Abdallah*[1], Anas Al-Rahamneh[2], Daniel Staegemann[3], Laila Barqawi[4]**

**Submitted**:13/03/2024 **Revised**: 28/04/2024 **Accepted**: 05/05/2024

**Abstract:** This research explores the intricate landscape arising from the integration of big data and large language models (LLMs) across sectors, unveiling intellectual property (IP) challenges requiring careful scrutiny. The transformative impact of big data and the ascendancy of LLMs in artificial intelligence have precipitated complex inquiries into data ownership, copyright law, and privacy. Central to these challenges is the ownership of datasets, especially those crucial in LLM training, reflecting the ambiguous nature of data as a contemporary digital asset. LLMs, proficient in generating content akin to their training materials, introduce nuances challenging traditional copyright boundaries. Privacy concerns escalate due to the pivotal role of personal data in both big data analytics and LLM functionality. This research aims for a comprehensive examination of these IP challenges by scrutinizing existing legal frameworks, evaluating their adequacy in the context of big data and LLMs, and unraveling the intricate relationship between technological innovation and IP law. The ultimate goal is to propose legal solutions or frameworks adept at tackling these emergent challenges. The significance of this research lies in its potential to shape robust legal and ethical standards in the digital age, providing valuable insights for policymakers, technologists, and legal experts to navigate the evolving nexus of technology and intellectual property.

***Keywords:*** *Intellectual Property, Big Data, Large Language Model, Artificial Intelligence.*

## 1. Introduction

In the evolving landscape of technology, two phenomena have emerged as particularly influential: big data and large language models (LLMs). The advent of big data, with its defining characteristics of volume, variety, and velocity, has revolutionized a myriad of sectors, enabling breakthroughs in analysis and decision-making [1, 2]. Alongside this, the rise of LLMs, which represent a significant stride in artificial intelligence (AI), has expanded our capabilities in natural language processing, leading to versatile applications [3].

However, this rapid integration of big data and LLMs into various domains has given rise to numerous intellectual property (IP) challenges. A primary concern is data ownership, a complex issue in the digital era where data is often likened to a valuable commodity. The ambiguity surrounding the ownership of datasets, particularly those used to train LLMs, poses a legal conundrum [3].

Copyright law adds another layer to these challenges. LLMs are capable of producing content that may closely mirror their training materials. This phenomenon raises questions about the originality of such outputs and their classification as derivative works, potentially infringing existing copyrights. The distinction between original creation and infringement is thus increasingly blurred in the context of LLMs [4, 5].

Privacy concerns are also paramount, as the utilization of personal data is integral to both big data analytics and LLM functionality. This raises issues regarding the alignment of technological advancements with privacy rights and regulations [6].

This research aims to investigate these IP challenges posed by big data and LLMs. The study will analyse existing legal frameworks, identify their limitations in the context of these technologies, and explore the interaction between technological innovation and IP law. The ultimate goal is to propose potential legal solutions or frameworks that address these emerging challenges.

The significance of this research lies in its potential to guide the development of legal and ethical standards in the digital age. As technology continues to advance, establishing robust legal frameworks becomes crucial. This study aims to contribute to the discourse on IP law in the realm of big data and LLMs, providing insights for policymakers, technologists, and legal experts.

## 2. Literature Review

Traditional intellectual property (IP) laws, established long before the advent of advanced artificial intelligence (AI), have interacted with AI technologies in a complex and often challenging manner. Historically, IP laws aimed to protect human creativity and innovation, focusing on works and inventions directly attributable to human authors and

[1] *Al-Zaytoonah University of Jordan, Amman, Jordan*
*ORCID ID : 0000-0002-3643-0104*
[2] *Public University of Navarra, Pamplona, Spain*
*ORCID ID : 0000-0003-3554-6604*
[3] *Otto von Guericke University, Magdeburg, Germany*
*ORCID ID : 0000-0001-9957-1003*
[1] *Al-Zaytoonah University of Jordan, Amman, Jordan*
*ORCID ID : 0000-0003-0435-2591*
** Corresponding Author Email: m.abdallah@zuj.edu.jo*

inventors. However, the emergence of AI, capable of generating creative works and potentially inventive solutions, has stretched the boundaries of these traditional legal frameworks. The relevance of IP laws in AI's context lies in their role in safeguarding the interests of creators, investors, and users of AI technologies, while also addressing the unique challenges posed by AI-generated content and inventions. This interaction raises fundamental questions about authorship, ownership, and the nature of creativity and invention in the AI era [7].

The interaction of artificial intelligence (AI) with key legal frameworks in intellectual property (IP) law reveals a landscape where traditional concepts are continuously re-evaluated. Major international treaties like the Berne Convention and the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) form the cornerstone of this legal domain [8].

The Berne Convention [9], established to protect the rights of authors over their literary and artistic works, traditionally hinges on human authorship. The application of this convention to AI-generated works is challenging because the convention presupposes human creativity and originality. This raises questions about the protection of works where the creative input is significantly from AI. Can AI-generated works, lacking this human creative element, be protected under the Berne Convention? This question remains a subject of debate among legal scholars and practitioners.

The TRIPS Agreement [8], which sets down minimum standards for many forms of IP regulation as part of the international trading system, also faces challenges in the AI context. Issues arise particularly in the realm of patents, where AI's role as an inventor or a tool in the invention process tests existing definitions. The agreement’s provisions were framed with human inventors in mind, leaving a gray area when AI significantly contributes to or independently creates an invention.

At a national level, various court cases and legal interpretations have started to shape the IP landscape in relation to AI. For instance, in the United States, court cases like Alice Corp. v. CLS Bank International [10] have impacted the patentability of AI-related inventions, particularly software and abstract ideas. Similarly, the European Patent Office has faced dilemmas regarding AI's role in inventions, leading to discussions about revising patent law to better accommodate AI's growing influence.

These legal frameworks and court cases exemplify the ongoing struggle to adapt traditional IP laws to the realities of AI. They highlight the need for a nuanced approach that balances the protection of human creativity and the novel issues brought about by AI technologies. The evolving interpretation of these laws and their application to AI-related cases will significantly shape the future of IP in the digital age [11].

The debate over AI's role as a creator or inventor in the realm of intellectual property (IP) law is marked by a collision of traditional legal concepts with modern technological capabilities. Central to this debate is whether an AI system can be recognized as a 'creator' or 'inventor' in the legal sense. Traditional IP frameworks are predicated on human authorship and invention, which raises complex questions when AI autonomously generates works or contributes to inventions [12].

Scholarly opinions are divided [13]. Some argue that recognizing AI as a creator or inventor could incentivize innovation and acknowledge the realities of modern creative and inventive processes. Others caution that this could undermine the foundational principles of IP law, which are designed to protect human creativity and ingenuity.

Landmark cases and decisions have begun to address this issue. Notably, the European Patent Office (EPO) and the United States Patent and Trademark Office (USPTO) have both faced applications where AI systems were listed as inventors. In decisions that have been mirrored globally, both offices rejected these applications, reiterating that under current laws, inventors must be natural persons. These decisions highlight the legal community's current stance but also underscore the growing pressure to reevaluate and potentially update IP laws in light of AI's evolving capabilities and roles [14, 15].

The protection of AI-generated works under copyright law is a topic of evolving legal interpretation, varying significantly across jurisdictions. Traditionally, copyright law requires a work to originate from a human author to qualify for protection [16, 17]. This principle, based on the concept of original human creativity, poses a challenge when applied to AI, which can create works independently of direct human input. In many jurisdictions, such as the United States, the European Union and Jordan [14, 18], the current legal framework does not explicitly recognize AI as an author, leaving AI-generated works in a legal gray area. However, some countries are exploring legislative changes to address this gap. For example, the UK's Intellectual Property Office has initiated discussions on whether and how AI-generated works should be protected. This varying stance across jurisdictions reflects the ongoing struggle to adapt copyright laws to the digital age, balancing the need for protection against the ethos of original human creativity.

The relationship between Intellectual Property (IP) and big data has evolved significantly over the years, with several key milestones shaping this intersection. Figure 1 1ist a timeline outlining the development of this relationship.

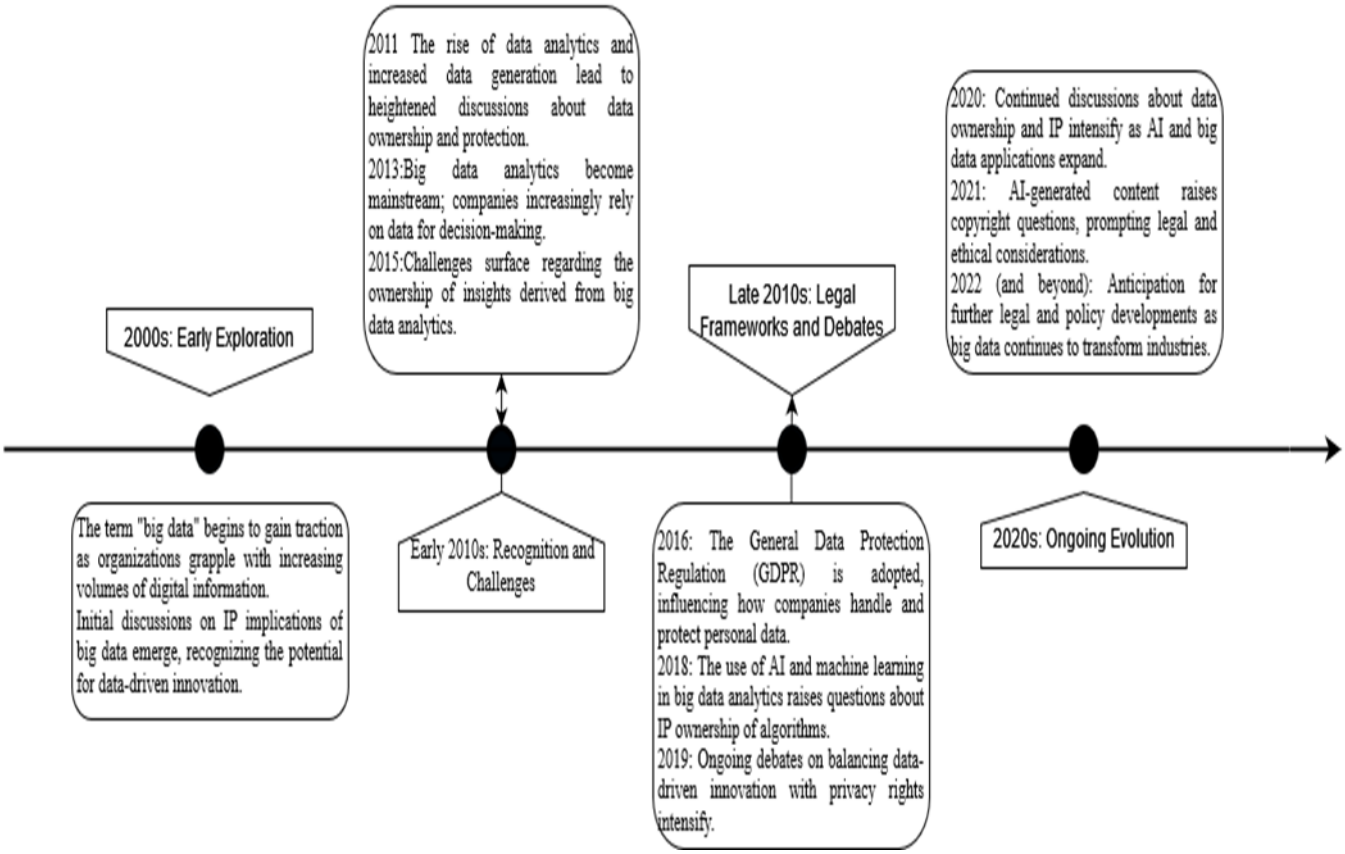


**Fig 1.** IP and Big data relation timeline

The timeline, reflects the evolving nature of the relationship between IP and big data, encompassing increased awareness, legal developments, and ongoing debates about ownership, privacy, and the ethical use of data-driven technologies. The trajectory is likely to continue as technology advances and society grapples with the challenges and opportunities presented by big data.

### 3. The Patentability of AI Technologies

Navigating the patentability of AI technologies entails challenges due to their unique nature. The United States Patent and Trademark Office (USPTO) offers guidelines reflecting a willingness to adapt, while The European Patent Office (EPO) maintains a stricter stance, necessitating a human inventor. This global disparity underscores the ongoing debate on whether existing patent criteria suffice for AI innovations or if new standards are imperative. These discussions are pivotal for shaping the future of IP protection for AI technologies. In this dynamic landscape, the established criteria of novelty, non-obviousness, and utility are applied, emphasizing the need to reassess assumptions in light of AI's distinctive inventive processes.

### 3.1. Criteria for Patenting AI Technologies

When considering AI technologies for patents, patent offices globally have had to navigate the unique characteristics of these inventions. USPTO has issued guidelines that offer some clarity on how AI-related inventions are evaluated, especially in terms of their novelty and non-obviousness. However, these guidelines still operate within the traditional patent framework, which does not specifically account for AI's role in the inventive process.

EPO has taken a more conservative stance. While AI-related inventions can be patented, the EPO maintains that AI systems cannot be recognized as inventors. This position was reinforced in decisions where the EPO refused patent applications naming an AI system as the inventor, indicating the importance of human contribution in the inventive process.

Current IP laws, designed primarily in the pre-digital era, often fall short in accommodating the nuances of AI and LLMs. These technologies present unique challenges that are not fully addressed by existing legislative frameworks.

One significant gap is the lack of clear guidelines on the attribution of inventorship in the context of AI. When an AI system plays a crucial role in the development of an invention, the traditional notion of 'inventor' becomes blurred. This raises questions about the ownership of AI-generated inventions and the distribution of rights associated with them.

Another area of concern is the treatment of data used by

LLMs and other AI technologies. Big data, which is essential for training and operating these systems, often comprises information from various sources, some of which may be protected under IP laws. The current legislation does not provide clear guidelines on how this data can be used without infringing on existing IP rights, creating a legal ambiguity.

The unique nature of LLMs and the scale of big data amplify these legislative gaps. LLMs, through their capacity to analyze and generate complex texts, raise questions about the originality and authorship of their outputs. This challenges the existing copyright frameworks which are grounded in human creativity.

Additionally, the scale and variety of big data used in training and operating AI systems pose significant challenges in IP law. Issues of data ownership, rights to access, and use of such data are not comprehensively addressed, creating potential conflicts and ambiguities.

Legal scholars and analysts have pointed out these gaps and called for reforms. Some advocate for the introduction of new categories of rights or revised definitions of existing IP concepts to better accommodate AI and big data [19]. Others suggest more radical changes, proposing entirely new frameworks to address the unique challenges posed by these technologies [20].

To put it briefly, while AI technologies offer immense potential for innovation, their integration into the current IP law framework is fraught with challenges. The patentability of AI inventions, the role of AI as an inventor, and the handling of big data in the context of IP law are areas that require significant reevaluation and legislative adaptation. The evolving nature of these technologies and their applications necessitate a dynamic legal response, one that balances the protection of traditional IP rights with the realities of the digital age.

### 3.2. Data Ownership and Privacy

Data ownership in big data and AI is a contentious and varied concept across jurisdictions. Data ownership and privacy are critical considerations in the evolving landscape of big data and artificial intelligence (AI). The legal definitions surrounding these concepts vary significantly across jurisdictions, shaping the way individuals and organizations interact with data.

In the United States, data ownership is often determined by the terms of service agreements and privacy policies set by service providers [21]. This approach can lead to the commercial exploitation of user data, with companies wielding considerable control over the information generated by users. This model raises concerns about transparency and user consent in the context of data usage.

On the other hand, the European Union has taken a distinct path with the General Data Protection Regulation (GDPR) [22]. This legislation empowers individuals by providing them with greater control over their personal data. Users have the right to know how their data is used, and companies are required to obtain explicit consent for data processing. The GDPR's user-centric approach aims to protect privacy rights in the digital age.

The divergence in legal frameworks has profound implications for the development and deployment of AI technologies globally. Striking a balance between fostering innovation and safeguarding individual privacy remains a complex challenge in the dynamic landscape of data ownership and privacy regulation.

The impact of privacy laws on intellectual property (IP) rights [23, 24] is pronounced in big data, where the tension between protecting creators' rights and safeguarding personal data becomes apparent. Intellectual property laws traditionally emphasize the protection of innovative creations, providing a framework for inventors and creators to safeguard their ideas and products. On the other hand, privacy laws are designed to ensure the confidentiality and security of personal data, placing restrictions on how information is collected, processed, and utilized.

In sectors heavily reliant on artificial intelligence (AI), such as machine learning and data analytics, personal data often serves as a crucial resource for innovation [25]. This intersection of IP and privacy laws introduces complexities and potential conflicts. Innovators may find themselves navigating a delicate balance between the need to protect intellectual property and the obligation to respect individuals' privacy rights.

The evolving landscape of technology and data-driven industries underscores the necessity of harmonizing these legal frameworks. Striking a balance that facilitates innovation while preserving individual privacy remains an ongoing challenge, requiring careful consideration of the implications for both creators and data subjects in the dynamic landscape of big data and AI.

Copyright law, a pillar of intellectual property protection, faces a paradigm shift in the era of Language Model Models (LLMs). Traditionally, copyright has been contingent upon human authorship, but the advent of LLMs challenges this fundamental premise. These models, such as OpenAI's GPT-4, have the ability to autonomously generate content, sparking a reevaluation of copyright principles [26].

At its core, copyright grants exclusive rights to the creator of an original work, allowing them to control the reproduction, distribution, and public display of their creation. However, when content is produced by LLMs without direct human input, the question arises: who holds the copyright?

The U.S. Copyright Office has taken a cautious approach, refusing to register works created by AI systems on the grounds of lacking human authorship [27]. This stance aligns with traditional copyright principles, which have historically required a human author to qualify for protection. One notable case involved a dispute over the copyright status of a computer-generated artwork known as "Edmond de Belamy," created by an algorithm. The U.S. Copyright Office's refusal to grant copyright registration to the artwork highlighted the legal system's struggle to adapt to the nuances of AI-generated content.

The issue becomes more complex when considering derivative works created by LLMs. Derivative works, modifications or adaptations of existing copyrighted material, have long been subject to copyright law. However, when LLMs autonomously generate content, determining the originality and authorship of subsequent works poses a considerable challenge.

The legal landscape is still evolving, and the lack of clear precedents complicates matters. Some argue that LLMs should be considered tools rather than authors, and therefore, the human operator using the model should be the rightful copyright holder. Others contend that the sheer autonomy and complexity of LLMs warrant recognition as quasi-authors, deserving some form of legal protection.

The ongoing debate surrounding LLMs and copyright law is not confined to the United States. Internationally, jurisdictions are grappling with similar questions. The European Union, for instance, is exploring the implications of AI on intellectual property, recognizing the need for a comprehensive and harmonized approach [28].

As LLMs continue to advance in sophistication, the legal system faces the challenge of adapting copyright law to accommodate these technological developments. The root of the matter lies in striking a balance between fostering innovation and creativity while ensuring fair and equitable protection for creators, whether human or machine.

The evolving nature of LLMs necessitates a nuanced understanding of their capabilities and limitations. These models are trained on vast datasets, learning patterns and generating content based on the information they have absorbed. While they lack consciousness and intentionality, their ability to produce human-like text blurs the lines between machine and human creativity.

One potential solution involves revisiting copyright law and considering a new category or framework that acknowledges the unique nature of AI-generated content. This might involve introducing a system where humans and AI collaboratively share copyright, recognizing the input and guidance provided by human operators.

### 3.3. Fair Use Considerations

The fair use doctrine plays a pivotal role in the ethical and legal considerations surrounding the training of Language Model Models (LLMs) using copyrighted material. In the United States, fair use is evaluated based on factors such as the purpose of use, the nature of the copyrighted work, the amount used, and the effect on the market value of the original work. However, when it comes to the massive scale at which LLMs consume and learn from copyrighted material during their training, the application of fair use becomes a complex and uncertain terrain [29].

The extensive utilization of diverse copyrighted content to train LLMs raises questions about the transformative nature of such use, as well as its potential impact on the market for the original works. The legal community grapples with defining the boundaries of fair use in the context of training LLMs, recognizing the tension between fostering innovation through machine learning and respecting the rights of copyright holders. As technology continues to advance, finding a balance that upholds fair use principles while addressing the challenges posed by LLM training practices remains a dynamic and evolving legal challenge.

## 4. International perspectives

International perspectives on intellectual property (IP) laws, particularly concerning artificial intelligence (AI) development, vary considerably among regions such as the European Union (EU), the United States, and Asia particularly China [30, 31].

The EU, particularly through instruments like the General Data Protection Regulation (GDPR), emphasizes individual privacy and data protection. This focus impacts how AI technologies, especially those reliant on big data, are developed and deployed. In terms of copyright, the EU does not explicitly recognize AI as an author or creator, which influences the protection available for AI-generated content. Regarding patents, the European Patent Office (EPO) maintains that inventors must be natural persons, thus excluding AI as an inventor.

The US has taken a more flexible approach to AI in IP law. The United States Patent and Trademark Office (USPTO) has been actively engaging with the AI community to understand the implications of AI in patent law and has issued guidelines for patenting AI-related inventions. However, similar to the EU, the US does not recognize AI as an inventor. In copyright, there is also no clear provision for AI-generated works, though the debate is ongoing, with some advocating for more adaptable interpretations.

China’s approach to AI and IP has been characterized by a rapid advancement and a strong push for innovation. The Chinese government has implemented policies that encourage AI development, reflecting in its IP law approach. China’s patent law has been more accommodating of AI-related inventions, and there are

discussions about further reforms to address AI's role in invention more explicitly. However, issues of data privacy and protection are not as stringently addressed as in the EU.

The major differences in these jurisdictions lie in their balance between promoting innovation and protecting traditional IP rights. The EU's stringent data protection laws, while safeguarding privacy, might limit certain AI innovations. The US, with its more flexible approach, encourages innovation but faces challenges in adapting its IP laws to AI's unique characteristics. China's strong focus on AI innovation demonstrates a different balance, prioritizing technological advancement, potentially at the expense of individual privacy rights and traditional IP norms.

These differences have significant implications for international companies and researchers working in AI. Navigating these diverse legal landscapes requires careful strategy and awareness of the varying legal risks and requirements in different jurisdictions. Moreover, these disparities underscore the need for international dialogue and potential harmonization of laws to address the global nature of AI and big data technologies.

The international variability in IP protection for AI highlights the dynamic and evolving nature of this legal field. The comparison of different jurisdictions reveals a spectrum of approaches, each with its implications for AI development and global collaboration. As AI continues to advance, it becomes increasingly important for IP laws to adapt, ensuring they adequately address the unique challenges posed by these technologies while fostering a conducive environment for innovation and protection of rights.

## 5. Case Studies of International Handling of AI and IP Issues

The international landscape of AI and IP law has been marked by several notable case studies, each reflecting the unique legal and cultural contexts of different countries.

### 5.1. Case Study 1: United States - Thaler v. Hirshfeld [32, 33]

In the landmark case of Thaler v. Hirshfeld in the United States, Dr. Stephen Thaler's patent applications, which listed an AI system named DABUS (Device for the Autonomous Bootstrapping of Unified Sentience) as the inventor, brought to the fore the issue of AI inventorship under U.S. law. The USPTO's rejection of these applications was grounded in the principle that under current U.S. patent law, only natural persons can be recognized as inventors. This decision was pivotal in highlighting the limitations of existing legal frameworks when confronted with the advanced capabilities of AI. The case not only sparked a national conversation about potential reforms in patent law to accommodate AI as an inventor but also influenced international discourse on this emerging issue. Thaler's subsequent appeal and the ongoing legal debate emphasize the growing need to reassess and potentially update IP laws in the age of AI.

### 5.2. Case Study 2: European Union - EPO's Refusal of AI Inventor Applications [34, 35]

The European Patent Office (EPO) faced a significant test with the submission of patent applications naming an AI system as the inventor. These applications, which also stemmed from Dr. Stephen Thaler's work with the AI system DABUS, were refused on the basis that under the European Patent Convention (EPC), an inventor must be a human being. This decision by the EPO was a clear reaffirmation of the traditional interpretation of inventorship in European IP law. However, it also marked a critical point in the broader discussion about the adequacy of existing IP frameworks in the face of rapidly advancing AI technologies. The EPO's stance influenced similar decisions in other jurisdictions and highlighted the divergences in international IP law regarding AI. The case underscores the complexities and legal challenges in recognizing AI's role in the creative and inventive processes within the current European legal framework.

### 5.3. Case Study 3: United Kingdom - The Copyright and AI Conversation [36, 37]

In the United Kingdom, the Intellectual Property Office initiated a significant conversation about the implications of AI-generated works for copyright law. This proactive approach was driven by the growing recognition of AI's capacity to create content that could potentially qualify for copyright protection. The UK's consultation process sought inputs from various stakeholders, including legal experts, AI developers, and the creative industry, to understand the challenges and opportunities presented by AI in the realm of copyright. This initiative reflected a willingness to adapt and evolve legal frameworks in response to technological advancements. The discussions centered on issues such as the ownership of AI-generated works, the application of existing copyright doctrines to such works, and the potential need for new legislative measures. The UK's engagement in this dialogue exemplifies the active role governments can play in shaping the legal landscape to address the novel challenges posed by AI in the field of intellectual property.

## 6. Global Trends And Consensus

The global approach to AI and IP law is characterized more by divergence than consensus. Different countries have taken varied stances on key issues like AI inventorship and copyright for AI-generated works, reflecting their legal traditions and innovation policies.

International organizations like the World Intellectual

Property Organization (WIPO) play a crucial role in shaping a cohesive approach to AI and IP law. WIPO has initiated discussions and research on the intellectual property implications of AI, aiming to understand the diverse perspectives of its member states. Through forums and consultations, WIPO seeks to facilitate international dialogue and potentially harmonize aspects of IP law in the context of AI [38].

Despite these efforts, there is currently no emerging global consensus on AI and IP law. The complexity of AI technology, combined with the varied legal, economic, and cultural contexts of different countries, contributes to continued divergence in legal approaches. While some countries emphasize innovation and adaptability, others prioritize traditional IP principles and the protection of human creativity [39].

The future trajectory of AI and IP law will likely involve a gradual convergence of some aspects, particularly around shared interests like fostering innovation and protecting data privacy. However, complete harmonization remains challenging due to the fundamental differences in legal philosophies and innovation strategies. The ongoing efforts of international organizations and cross-border collaborations will be vital in navigating these complexities and shaping a more unified global approach to AI and IP law.

## 7. Challenges and Proposed Solutions

The intersection of artificial intelligence (AI) and legal frameworks presents several challenges that demand careful consideration and innovative solutions. One of the primary challenges is the ambiguous legal status of AI-generated content. Traditional copyright laws often require human authorship, leaving AI-generated works in a legal gray area. Addressing this challenge involves updating copyright laws to explicitly recognize and accommodate the unique nature of AI-generated creations. This may include establishing criteria for authorship and ownership in cases where AI plays a significant role in content creation.

Another complex issue arises from the conflict between data privacy and intellectual property (IP) rights. As AI systems rely heavily on vast datasets, navigating the balance between protecting personal data and fostering innovation becomes crucial. Solutions to this challenge could involve creating guidelines that delineate the ethical and legal boundaries for data usage in AI development. Establishing clear consent mechanisms, data anonymization practices, and transparency requirements can contribute to building trust between AI developers and data subjects.

The international variability in legal frameworks adds an additional layer of complexity. Harmonizing IP laws globally can create a more cohesive framework for AI development, ensuring consistency and predictability across borders. This may involve international collaboration to establish standards and principles that accommodate diverse cultural and legal perspectives while fostering a shared understanding of the ethical and legal considerations surrounding AI.

Proposed solutions also include the development of specific guidelines for AI development that balance data privacy concerns with the necessity of using data in training models. Ethical AI guidelines could help mitigate risks associated with biased algorithms and unauthorized use of personal information. Such guidelines could be crafted collaboratively by experts, policymakers, and industry stakeholders to ensure a comprehensive and responsible approach to AI development.

Additionally, public awareness and education play a vital role in addressing these challenges. Informing the public about AI technologies, their implications, and the legal safeguards in place can empower individuals to make informed decisions regarding their data and contribute to a more ethical and responsible AI ecosystem.

## 8. Conclusions

This research delves into the complex intersection of big data and large language models (LLMs), unraveling intricate intellectual property (IP) challenges that have emerged with the transformative impact of these technologies. The ownership of datasets, particularly in LLM training, presents a central challenge, highlighting the ambiguous nature of data in the digital era. LLMs, capable of mimicking training materials, introduce nuances challenging traditional copyright boundaries, while escalating privacy concerns necessitate careful scrutiny. The research advocates for a comprehensive examination of these IP challenges, assessing existing legal frameworks, and proposing adaptive solutions. Its significance lies in shaping robust legal and ethical standards for policymakers, technologists, and legal experts, offering guidance in navigating the evolving landscape of technology and intellectual property in the digital age.

## Author contributions

All authors contribute equally in this research.

## Conflicts of interest

The authors declare no conflicts of interest.

## References

[1] M. Abdallah and M. Salah, "Artificial Intelligence and Intellectual properties: legal and ethical considerations," International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 1, pp. 368-376, 2024.

[2] L. Barqawi and M. Abdallah, "Copyright and generative AI," Journal of Infrastructure, Policy and

Development, vol. 8, no. 8, p. 6253, 2024.

[3] E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," Learning and Individual Differences, vol. 103, p. 102274, 2023/04/01/ 2023, doi: https://doi.org/10.1016/j.lindif.2023.102274.

[4] K. Taylor, "The patentability of inventions with artificial intelligence listed as an inventor following Thaler v. Hirshfeld," IPCLJ, vol. 6, p. 1, 2021.

[5] M. Abdallah, K. Jaber, M. Salah, and A. Omayah, "Impact of Big Data Analytics and Large Language Model on Intellectual Property," Al-Zaytoonah University of Jordan Journal for Legal studies, vol. 5, no. Special Issue, pp. 1039-1047, 2024.

[6] [S. Ghayyur, J. Averitt, E. Lin, E. Wallace, A. Deshpande, and H. Luthi, "Panel: Privacy Challenges and Opportunities in {LLM-Based} Chatbot Applications," 2023.

[7] S. Tripathi and C. Ghatak, "Artificial intelligence and intellectual property law," Christ ULJ, vol. 7, p. 83, 2018.

[8] C. M. Correa and A. A. Yusuf, Intellectual Property and International Trade: The TRIPS Agreement: The TRIPS Agreement. Kluwer Law International BV, 2016.

[9] S. Ricketson and J. Ginsburg, "International copyright and neighbouring rights: The Berne Convention and beyond," 2022.

[10] J. Clizer, "Exploring the Abstract: Patent Eligibility Post Alice Corp v. CLS Bank," Mo. L. Rev., vol. 80, p. 537, 2015.

[11] C. Schmit, M. Doerr, and J. Wagner, "Leveraging IP for AI governance," Science, vol. 379, no. 6633, pp. 646-648, 2023.

[12] L. Barqawi and S. Al-Arasi, "Intellectual Property Protection through University Innovation Centres: Jordan's Approach to Sustainable Development," International Journal of Religion, vol. 5, no. 5, pp. 542-548, 2024.

[13] I. Abdikhakimov, "Legal aspects of AI generated content," in International Conference on Legal Sciences, 2023, vol. 1, no. 5, pp. 1-17.

[14] P. Jandrić, "On The Hyping of Scholarly Research (With A Shout-Out to ChatGPT)," Postdigital Science and Education, pp. 1-8, 2023.

[15] C. G. Merritt, "A Compulsory Solution to the Machine Problem: Recognizing Artificial Intelligence as Inventors in Patent Law," Vand. J. Ent. & Tech. L., vol. 25, p. 211, 2023.

[16] P. Cummings, A. Balazer, and C. Worrel, "Who is an Inventor?," 2023.

[17] K. Hristov, "Artificial intelligence and the copyright dilemma," Idea, vol. 57, p. 431, 2016.

[18] O. Taha, "Intellectual Property Rights in Jordanian Legislation A Comparative Study with Islamic Jurisprudence a New Endowment Formula " Al-Zaytoonah University of Jordan Journal for Legal studies, vol. 2, no. 3, pp. 133-153, 2021, doi: 10.15849/ZUJJLS.211130.06.

[19] B. J. Keele and H. Alcasid, "Copyright and Non-Humans," 2018.

[20] J. Oster, "Code is code and law is law—the law of digitalization and the digitalization of law," International Journal of Law and Information Technology, vol. 29, no. 2, pp. 101-117, 2021.

[21] J. Chen et al., "Copy, right? a testing framework for copyright protection of deep learning models," in 2022 IEEE Symposium on Security and Privacy (SP), 2022: IEEE, pp. 824-841.

[22] T. D. Haley, "Illusory Privacy," Ind. LJ, vol. 98, p. 75, 2022.

[23] S. Unit, "The General Data Protection Regulation (GDPR)," 2022.

[24] C. Radsch, "Weaponizing Privacy and Copyright Law For Censorship," CIGI Papers No, 2023.

[25] L. Barqawi, "The access to medicine puzzle: scaling back the negative effects of the Jordan–US Free Trade Agreement," Journal of Intellectual Property Law & Practice, vol. 14, no. 9, pp. 678-686, 2019, doi: 10.1093/jiplp/jpz080.

[26] J. Luo, "Data-driven innovation: What is it?," IEEE Transactions on Engineering Management, vol. 70, no. 2, pp. 784-790, 2022.

[27] P. Samuelson, "Generative AI meets copyright," Science, vol. 381, no. 6654, pp. 158-161, 2023.

[28] P. Mezei, "'You AIn't seen nothing yet': Arguments against the protectability of AI-generated outputs by copyright law," in Law, Regulation and Governance in the Information Society: Routledge, 2022, pp. 126-143.

[29] T. Awad, "Universalizing Copyright Fair Use: To Copy, or Not to Copy?," J. Intell. Prop. L., vol. 30, p. 1, 2022.

[30] L. A. DiMatteo, C. Poncibò, and M. Cannarsa, The Cambridge handbook of artificial intelligence: global perspectives on law and ethics. Cambridge University Press, 2022.

[31] K. Muehlfeld and M. Wang, "Intellectual Property Rights in China—A Literature Review on the Public's Perspective," Frontiers in Sociology, vol. 7, p. 793165, 2022.

[32] M. T. Messina, "Thaler v. Vidal, 43 F. 4th 1207 (Fed. Cir. 2022)," DePaul J. Art Tech. & Intell. Prop. L, vol. 33, p. 67, 2023.

[33] M. Lavrichenko, "Thaler v. Vidal: Artificial Intelligence-Can the Invented Become the Inventor," Cardozo L. Rev., vol. 44, p. 699, 2022.

[34] I. Buzu, "The DABUS Affair," Available at SSRN 4175009, 2022.

[35] R. Moufang, "Artificial Intelligence and the Technicality Requirement of Patent Law," in A Critical Mind: Hanns Ullrich's Footprint in Internal Market Law, Antitrust and Intellectual Property: Springer, 2023, pp. 471-488.

[36] A. Drake et al., "Legal contestation of artificial intelligence-related decision-making in the United Kingdom: reflections for policy," International Review of Law, Computers & Technology, vol. 36, no. 2, pp. 251-285, 2022.

[37] E. Bonadio, L. McDonagh, and P. Dinev, "Artificial intelligence as inventor: exploring the consequences for patent law," Intellectual Property Quarterly, vol. 1, pp. 48-66, 2021.

[38] S. Kokane, "The Intellectual Property Rights of Artificial Intelligence-based Inventions," Journal of Scientific Research, vol. 65, no. 2, 2021.

[39] K. Wills, "Ai around the world: Intellectual property law considerations and beyond," J. Pat. & Trademark Off. Soc'y, vol. 102, p. 186, 2021.