



The state of AI and intellectual property – a thematic scientometric assessment

Gergely Ferenc Lendvai¹  · Peter Mezei²  · Anett Pogacsas³ 

Received: 4 August 2025 / Accepted: 7 May 2026
© The Author(s) 2026

Abstract

This study presents a comprehensive scientometric analysis of the scholarly discourse on artificial intelligence and intellectual property law between 2013 and 2023. Drawing on metadata from Scopus, the research employs advanced bibliometric tools to map trends, publication patterns, and thematic clusters in the field. Our results reveal a dramatic surge in academic output post-2019, showing an accelerated scholarly interest, especially in light of generative AI's emergence. The CiteSpace-based analysis identifies three dominant research clusters: (1) AI and Copyright Law, focusing on authorship, originality, and text/data mining; (2) AI and Digital Rights, exploring identity, privacy, and ethical dilemmas; and (3) AI and Patent Law addressing inventorship, patentability, and system reform. Furthermore, we also expose geographical disparities, with the United States and United Kingdom dominating the field of AI and IP law, alongside severe under-representation of Global South-affiliated authors within Scopus-indexed, English-language journal publications. A key finding is that despite growing regulatory responses, such as the EU AI Act and US Copyright Office guidelines, legal consensus remains elusive. In sum, the article at hand aims to offer a state-of-the-field mapping of Scopus-indexed scholarship at the AI-IP interface via consolidating the field's "visible contours," while identifying underexplored thematic and geographic areas for future interdisciplinary inquiry.

Keywords Artificial intelligence · Intellectual property law · Bibliometric assessment · Copyright · Patent · Digital rights

✉ Gergely Ferenc Lendvai
lendvai.gergely.ferenc@uni-nke.hu

¹ Ludovika University of Public Service, Budapest, Hungary

² University of Szeged, Szeged, Hungary & Vytautas Magnus University, Kaunas, Lithuania

³ Pazmany Peter Catholic University Faculty of Law and Political Sciences, Budapest, Hungary

Introduction

Academic research on the trends and effects of artificial intelligence (AI) on law and especially intellectual property (IP) law has gradually become one of the trendiest “buzzes” on a global scale. Its relevance has been accelerated by the relatively recent appearance of generative AI (GenAI) models and applications and its importance was further exacerbated by the growing tensions surrounding access to training data as well as geopolitical conflicts related to access to hardware (“chip war”) or training data “distillation” (the Chinese DeepSeek was accused of training its R1 model on data collected from, or “distilled” via, OpenAI’s ChatGPT service).

A great bunch of legal concerns could be easily named. These include, first, the *technological nuances* of the input phase (the training of AI or machine learning) and the output phase (the generation of synthetic contents) (Chiou 2020; Schafer et al. 2015; Mylly 2023; Picht and Thouvenin, 2023; Hugenholtz and Quintais 2021). Second, *substantive and fundamental IP questions* are scrutinized and revisited on a daily basis, like the concept authorship or inventorship; the threshold of originality/novelty; the scope of rights and their limitations; the expected remuneration for the use of training data or content generation; and the direct or indirect liability for the use of AI (Senfleben 2023; Kahveci 2023; Stockton-Brown 2023; Rudzite-Celmina 2023;). Third, *procedural issues*, like registration, evidencing and jurisdiction were addressed by numerous commentators (Weber-Wulff et al. 2023; Qiao et al. 2023). Fourth, the empirical analysis of *contractual stipulations* has rapidly grown, and led to the quantitative and qualitative analyses of how end-user license agreements affect the use of the relevant GenAI models (Chaudhary 2023; Lone and Naaz 2021; Khan et al. 2021; Erb 2022). Fifth, the review of the *social costs* of whichever legal solution is preferred has gained momentum, too (Elkin-Koren 2020). Numerous authors (Mammen and Richey 2020; Bonadio and Mcdonagh 2020; Kalpokiene and Kalpokas 2023) set the existential threat of a pro-AI approach for incumbent human creators and the overall industry against the potential positive externalities a pro-AI environment could trigger. Finally, the *race to regulate* (or occasionally not to regulate) the development, the use and the effects of (Gen)AI has sparked intensive debates (Sganga 2023; Li et al. 2023; Zyhrii et al. 2023; Skvortsova et al. 2023; Wach et al. 2023).

Keeping ourselves up to date in the mass of research outputs, news reports, legislative changes and case law of this field has become burdensome, at best, and a mission impossible, realistically. Therefore, the need to have a clustered bibliometric assessment of the main avenues of this field has become urgent. The aim of this paper is exactly to understand the focal points of recent AI and IP research, to highlight the internal trends of this field, including the overlaps of academic papers, shared references (bibliographic coupling), keyword co-occurrence, or country authorship. Based on this research and analysis, the issues under-researched also seem to be clearly outlined, which may be important insights for further research.

To study the current state and possible research venues in the future, we propose two research questions (RQ):

- 1) *What are the current publication trends in the field of AI and IP law?*

2) *What are the key themes in the field?*

Via addressing these two questions, we position the paper as a state-of-the-field empirical mapping of scholarly developments at the intersection of AI and intellectual property law. Given the nature of bibliometric studies, our contribution is primarily descriptive, however, we also set forth to consolidate dispersed debates, visualize their internal organization, and highlight patterns of thematic concentration as well as the geographic distribution of scholarship. Where doctrinal or policy implications are discussed, these are analytically distinguished from the underlying scientometric findings.

Materials and methods

For the present research, we used Elsevier's Scopus database, one of the comprehensive and renowned scientific repositories that stores highly reliable metadata on both journals and publications (Baas et al. 2020; Zhu and Liu 2020). As IP law and its connection to AI are highly connected to social sciences, the use of Scopus is also supported by the fact that its coverage of social sciences-related publications is more inclusive and broader than other databases, such as the Web of Science (Norris and Oppenheim, 2007; Demeter 2020).

Data collection

We employed well-established bibliometric methods to collect data on publications (Donthu et al. 2021). For the data collection, we created a specific search string in Scopus to include a wide range of research endeavours related to IP law and AI:

TITLE-ABS-KEY("artificial intelligence" OR "AI" OR "machine learning" OR "deep learning" OR "neural networks" OR "natural language processing" OR "NLP" OR "generative AI" OR "generative models" OR "AI originality" OR "AI-assisted" OR "synthetic data") AND TITLE-ABS-KEY("intellectual property" OR "IP law" OR "copyright" OR "trademark" OR "patent" OR "trade secrets" OR "intellectual property rights" OR "IPR" OR "data ownership" OR "AI-generated content" OR "IP policy" OR "AI authorship" OR "AI inventorship" OR "fair use" OR "licensing" OR "open access" OR "creative commons") AND PUBYEAR > 2012 AND PUBYEAR < 2024 AND (LIMIT-TO (SRCTYPE, "j")) AND (LIMIT-TO (PUBSTAGE, "final")) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (SUBJAREA, "SOCI"))

As seen in the search string, we employed multiple restrictions to avoid filtering out papers irrelevant to our research. Firstly, as IP law is not a homogenous field of legal study (Calboli 2021), we included keywords that directly relate to issues that emerge in connection to AI. For this, we employed keywords such as "synthetic data" and "AI-assisted" to cover questions that are concerning the intersection of IP law issues and artificially generated content (see Offenhuber 2024).

We restricted the framework to an 11-year period between 2013 to 2023. The reason behind this consideration stems from the fact that we conducted preliminary

research on literature before 2013 and found that the topic IP law and AI is largely undiscussed mainly due to the lack of technological developments back then. 2023 was chosen as the last full year available at the time of the writing of the present article and also because of the fact that journals – primarily in the field of social sciences – publishing papers in this subject tend to follow slower procedures which may result in delays in publications (Björk and Solomon 2013). Similarly to other bibliometric research, we limited the scope to articles as these article types are less prone to publication bias, are subjected to more rigorous peer reviewing processes in general, and provide metadata that make them ideal for systematic scientometric analyses as opposed to review papers and book chapters (Manterola et al. 2013; Grant and Booth 2009; Munn et al. 2018; Levine et al. 2004). Lastly, we limited subject categories to social sciences as this category best fits our research goals. As a note, we shall also acknowledge that several conceptual debates relevant to the present analysis predate the widespread emergence of contemporary artificial intelligence technologies. In particular, earlier scholarship – some dating back to decades – on computer-generated works addressed questions of authorship, originality, and legal attribution (see Samuelson 1986; Hart 1988; Miller 1993; Glasser 2001) that closely parallel issues now discussed in the context of AI-generated content (for a very early analysis, see Davies 2011). While this set of literature falls outside the temporal scope of the present dataset and therefore is not captured by the applied search strategy, it constitutes an important doctrinal and conceptual precursor to current debates at the intersection of AI and intellectual property law.

Following the initial retrieval, a manual screening step was applied to exclude 446 records that constituted false positives with respect to the research focus. All authors jointly screened the abstracts and made the decision on the inclusion and exclusion together. The exclusions primarily concerned publications in which references to “artificial intelligence” or “intellectual property” were used in non-legal, technical, or metaphorical contexts (e.g. AI discussed purely as an engineering or computational problem, or “intellectual property” used to denote technological assets or proprietary tools rather than a legal construct), and which therefore did not substantively engage with intellectual property law.

Regarding the specific exclusion strategy, despite the specificity of the search query, a number of false positives were identified that did not examine AI and IP law jointly at all. These included, for instance, empirical AI case studies that merely mentioned intellectual property in passing (such as references to patented sets of data or software ownership), studies focusing on AI detection or watermarking tools without any legal analysis, infrastructure- or hardware-oriented research (e.g. system design or optimization), or applied AI research in domains where IP (in general terms, not specifically IP law) was referenced only tangentially and without doctrinal or regulatory discussion. Including such records would have introduced thematic noise and distorted the subsequent keyword co-occurrence and clustering analyses by inflating connections unrelated to legal doctrine or regulation. The exclusion process was applied conservatively, with borderline cases retained, to preserve analytical validity while minimizing the risk of over-filtering relevant legal scholarship. (Fig. 1)

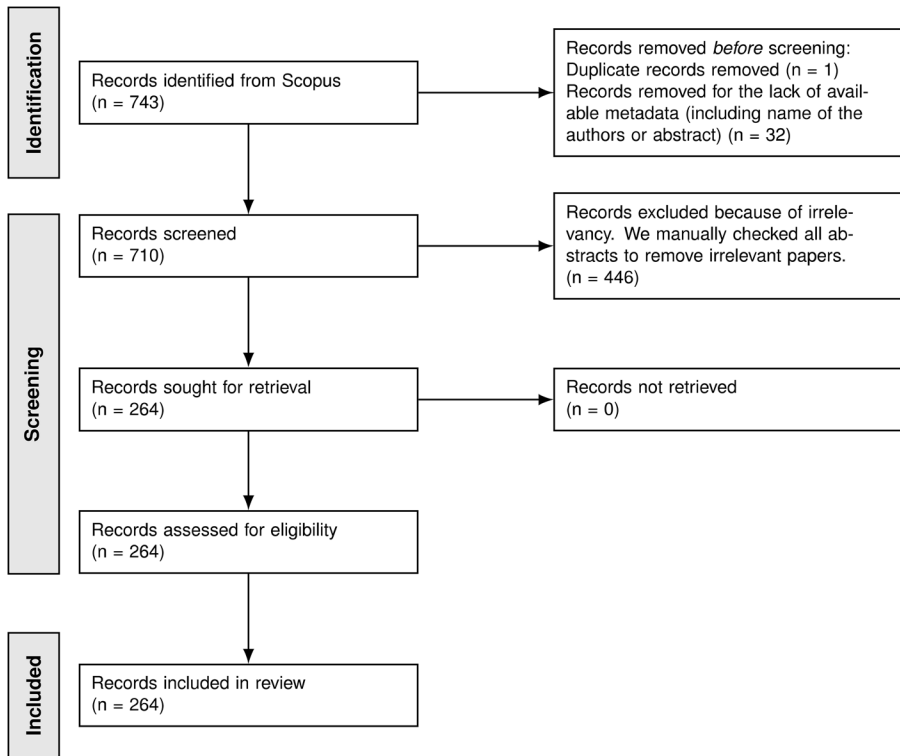


Fig. 1 Data collection procedure

Data analysis

We employed a multi-methodical review to analyze the data collected. For the descriptive bibliometric examination, we conducted publication count and citation analyses (*RQ1*) (Donthu et al. 2021; Meho 2007; Cristián et al. 2022). With regard to RQ1, following Cristián et al.'s (2022) and Lendvai's (2025) research on publishing practices related to legal issues, we were interested in the localization of knowledge produced from a geographical standpoint. To investigate this, we employed VOSviewer (Van Eck and Waltman 2017) to a country-based co-authorship network. In order to produce comprehensive results, we ignored papers with more than 25 authors, however, we did not limit the number of documents per author or country.

Following similar research endeavors on scientometric reviews related to highly novel issues such as new media phenomena or disinformation (Cataldo et al. 2022; Lendvai 2025; Tătaru et al. 2024), and to enrich the rather rudimentary research field of systematic reviews in the field of law (see Gupta et al. 2023), we conducted a topical analysis with CiteSpace (version 6.4.R2 Advanced), a software widely used in bibliometric analyses (Ding and Yang 2022) due to its extremely complex and comprehensive analytical tools and powerful visualizations (Chen 2006). Following Cataldo et al. (2022), we specified our analysis in terms of the g-index values ($k=125$; in order to prioritize highly referenced nodes and to discover influential

papers and prevalent keywords) and slice length (1) to cover papers year by year. The network data is as follows:

- Technical parameters related link pruning and clustering
 - Link retaining factor which controls the numbers of links retained, $LRF=2.5$
 - Link per node ratio which defines the number of links per node, $L/N=10$
 - Limitations to the number of backward years, $LBV=5$
 - Threshold for edge weight, $e=1.0$
- Network data:
 - Nodes (N)=682
 - Links/edges (E)=1956
 - Density=0.0084
 - Explanation: Our network contains 682 nodes and 1,956 links, with a very low density (0.0084), meaning it's a sparse network, which is typical for such networks.
 - Largest connected component, $n=1$, $CC=413$ (~61) meaning that the largest connected component – as in, the largest cluster of linked papers, contains 413 Ns making up 60% of the entire network.
 - Modularity, $Q=0.8957$ which underlines that our network has a very well-separated clustering structure with distinct, identifiable clusters and topics
 - Weighted Mean Silhouette, $S=0.9753$, which means that our network resulted in a very high silhouette score, showing excellent internal consistency within clusters.
 - Harmonic Mean, $Q, S=0.9338$ - The Q, S combines the aforementioned Q and S into one index; our value suggests that the network has both strong modularity and high cluster quality which is an excellent result for analysis.

With regard to clusters, we used CiteSpace's indicators (Chen 2006):

- Coverage: the number of members from the cluster that are cited by the article in question.
- GCS (Global Citation Score): The total number of times the article has been cited across Scopus.
- LCS (Local Citation Score): The number of times the article has been cited within the dataset we built.
- Citing articles: The number of articles within our dataset that cite the respective paper (Fig. 2).

All visualizations are the works of the authors. If not indicated otherwise, we used Python's matplotlib package to visualize data (Hunter 2007).

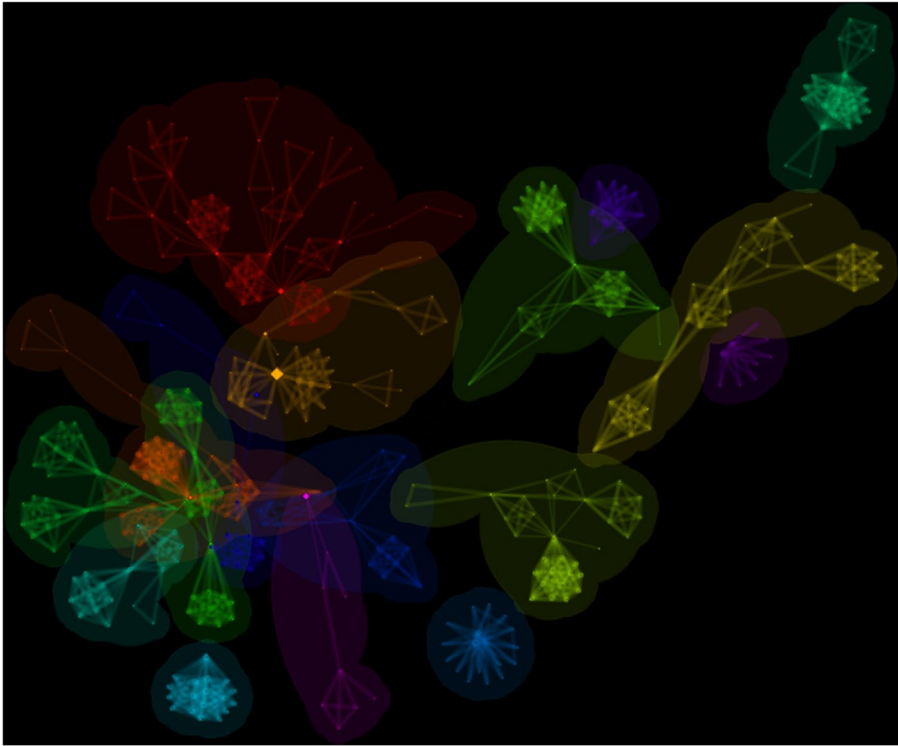


Fig. 2 Clusterized keyword network by degree of AI and IP law research. Each colored cluster represents a different set of research topic. For a detailed summary see Appendix 1

Results

The number of academic publications related to AI and IP law has shown a clear upward trend, especially after 2019. In 2013, only 2 publications were recorded, followed by a modest increase to 3 in 2014. The count remained relatively low and stable until 2017, fluctuating between 1 and 3 annually. However, from 2018 ($n=9$) onward, the field experienced noticeable growth, culminating in a peak of 100 publications in 2023. The results of a linear regression analysis support this trend ($R^2=0.73$, $p=0.00079$ ***). The mean publication/year was 24 during the examined period ($sd=31.41$, $min=1$, $25%=2$, $med=9$, $75%=39$, $max=100$) and publications are heavily concentrated in most recent years (Gini=0.62). (Fig. 3)

The citation distribution between 2013 and 2023 reveals substantial variation in impact across publication years. Early years such as 2015 and 2017 show high average citation counts (2017=19), though these are based on just one or two papers, indicating possible outlier effects. 2014, despite only having three publications, had a mean citation count of 12, suggesting those few studies gained moderate visibility. Conversely, more recent years like 2022 and 2023 exhibit lower median and mean citation values, which is expected due to their recency and shorter citation window. 2021 stands out with both a high number of publications and relatively high cita-

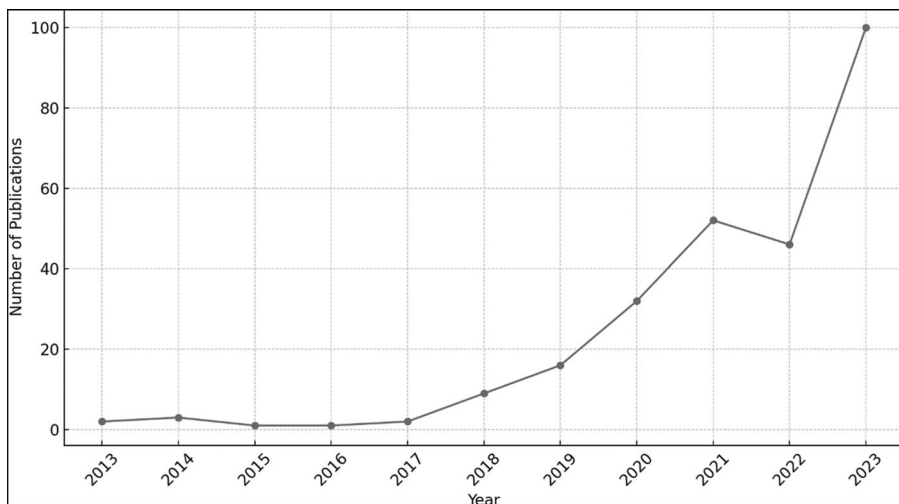


Fig. 3 Publications on IP law and AI between 2013 and 2023. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

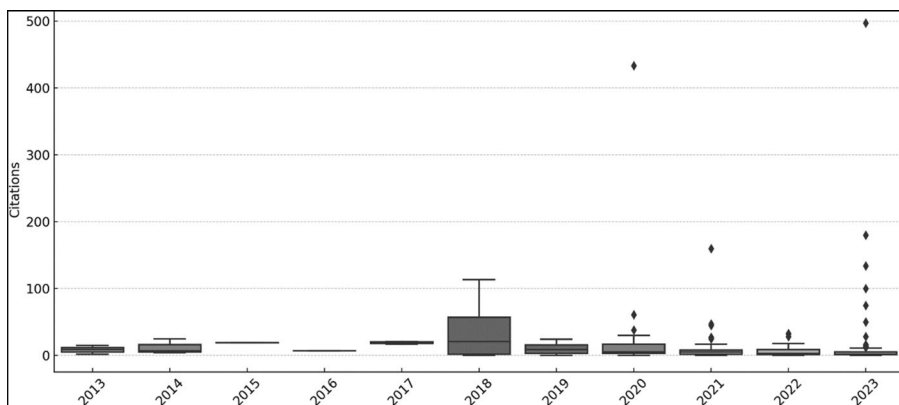


Fig. 4 Citation distribution between 2013 and 2023

tion dispersion. The standard deviation values indicate that some years (like 2020, $sd = 75.75$) had widely varying citation counts, pointing to a few standout publications skewing the average. The citation behavior confirms a common bibliometric trend: older publications have had more time to accumulate citations, while newer ones are still in the early stages of visibility (Higham et al. 2017). (Fig. 4)

Finally, we examined internationalization via a country-based co-authorship analysis using VOSviewer. Our network consists of 38 countries (items), organized into 8 clusters, with 73 co-authorship links and a total link strength of 86, indicating moderate but concentrated collaboration activity with an extreme Euro- and US-centric focus.

The United Kingdom (cluster 5, links = 19, total link strength = 29, documents = 33) emerges as one of the most central and highly connected nodes. The UK serves as a

“bridge” between multiple clusters including continental Europe, the United States, and parts of Asia. The United States (cluster 6, links=11, total link strength=14, documents=41), although slightly peripheral in placement, maintains strong bilateral ties – particularly with South Korea, France, and China – and serves as another major hub of collaboration (see Sasvári and Lendvai 2025). Germany (cluster 4, links=11, total link strength=16, documents=24) forms a dense subnetwork in collaboration with Sweden, Czech Republic, and Ukraine. China is also relatively central, with notable connections to Austria, Singapore, Spain, and Hong Kong. Countries like Netherlands, Italy, and France help anchor mid-size European clusters, while Australia (cluster 7, links=4, total link strength=6, documents=17) stands out as an isolated but clearly defined node with a few strong ties.

The co-authorship network revealed a noticeable under-representation of Global South (based on the affiliation of authors). Aside from a marginal appearance of a handful of African, Latin American, and Asian countries, most countries in these regions do not appear in the co-authorship network. This pattern is by no means new. Much rather, it reflects long-standing disparities in visibility and inclusion within major bibliometric databases, instead of the absence of relevant scholarship per se, and is consistent with broader structural inequalities in global academic knowledge production. Still, we must underline once more that this statement only concerns our specific data stemming from Scopus, thus, a general statement on the exclusion of Global South scholarship would be an overstatement. (Fig. 5)

Focusing on the core of our paper, the topical analysis, we identified 17 major clusters (see Fig. 2). However, as Cataldo et al. (2022), the automatic clustering of

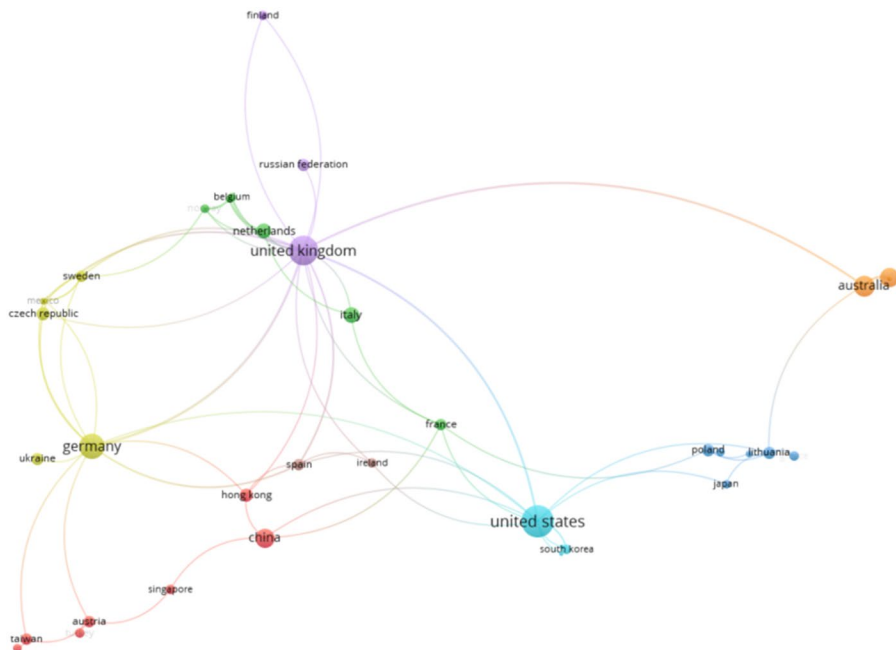


Fig. 5 Country-based co-authorship analysis

CiteSpace faces limitations in terms of the mentioned automatic denomination of the different clusters. Furthermore, since we base our research on keywords and not on the document co-citation analysis (DCA) employed by Cataldo et al. (2022), we manually checked the prominence of topics based on the degree (number of direct connections a node (keyword) has to others in the network, indicating how widely a concept is discussed across different topics), centrality, and sigma values, three well-established measures in scientometric network analysis.

For example, *copyrights* had the highest degree in the entire dataset (51), followed by *learning systems* (35), *artificial intelligence* (29), *commerce* (27), and *patents and inventions* (24), signalling that these terms are the most frequently and broadly connected within the field. Our second indicator, centrality, highlights a node's role as a bridge between different clusters, i.e., how crucial a concept is in linking otherwise separate discussions. Again, *copyrights* ranked highest (0.43), followed by *artificial intelligence* (0.28), *data ownership* (0.21), *social media* (0.21), and *patents and inventions* (0.09). Sigma, a composite indicator of centrality and burstiness, highlights topics that are both influential and emerging. Here too, *copyrights*, *artificial intelligence*, *data ownership*, *social media*, *innovation*, and *patents and inventions* all returned a sigma value of 1.00. As no significant citation bursts were detected across the dataset, sigma did not vary between nodes; it is therefore reported for completeness rather than as a discriminating indicator.

Based on this triangulated evaluation, we selected the following three most prominent thematic clusters for deeper analysis: (1) AI and Digital Rights, (2) AI and Copyright, and (3) AI and Patent Law. Nonetheless, we accentuate that these clusters should be understood as structurally prominent within the indexed dataset rather than as definitive indicators of doctrinal centrality across the broader field.

The topic of AI and Copyright emerged as a central cluster due to the unquestionable prominence of the term “copyrights”, which ranks first across all three measures. It also holds the highest degree (51), indicating the largest number of direct connections to other keywords, suggesting that copyright issues are discussed across a wide range of subtopics in AI and IP law. Moreover, this cluster also serves as a critical bridge connecting disparate scholarly discussions, such as authorship, content generation, text and data mining, and creative industries and has the highest centrality score (0.43) as mentioned earlier. Our second cluster, AI and Digital Rights was chosen based on the prominence of keywords addressing data governance, expression, privacy, and identity, core themes in the digital rights framework. The third cluster, AI and Patent Law, was selected on the basis of the keyword *patents and inventions*, which ranks fifth in degree (24), tenth in centrality (0.09), and holds a sigma value of 1. Although this topic has slightly lower centrality than others, its degree suggests it is a well-established and active field. (Fig. 6 and see Appendix 1: 18–19)

Discussion

The quantitative analysis showed a dramatic increase in publication output after 2019 which highlights that the field of AI and IP law has entered a phase of accelerated growth, closely aligned with the public and regulatory emergence of generative AI.

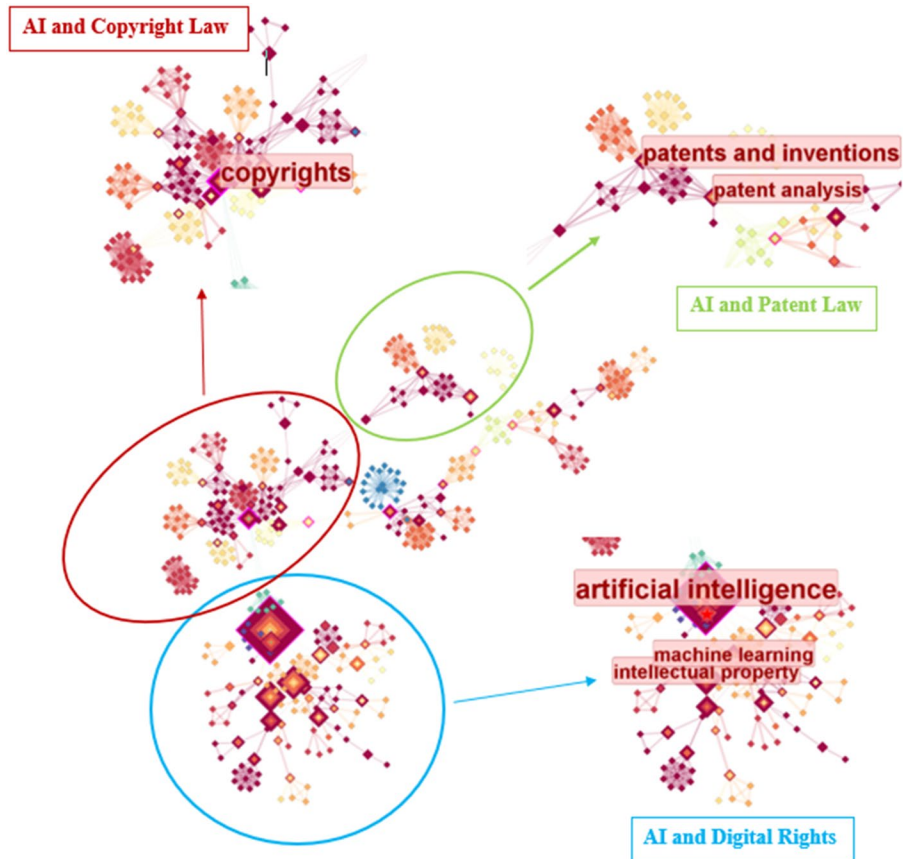


Fig. 6 Three major thematic clusters based on keyword selection by degree. The size of the nodes corresponds to the total n of occurrence, node labels are selected based on their top prominence using selective thresholds to only display the most important keywords

This pattern, however, has further implications. Not only does the growth trend demonstrate a rising sense of urgency around the legal implications of AI, but it also suggests a structural transformation in how IP-related legal research is being conducted and prioritized. Although earlier papers enjoy a citation advantage by virtue of time (Higham et al. 2017), the emergence of novel highly cited documents (such as several outliers published in 2023) shows a shifting (citation) dynamic in which more recent works gain visibility and citation traction faster within the dataset. However, we must underline that whether such prominence translates into long-term doctrinal “canonical” status cannot be inferred from citation-based indicators alone. In terms of global scholarly engagement, the country-based co-authorship analysis reveals a heavily uneven geographic distribution of research activity. Our findings fully confirm earlier endeavors pertaining to law-related scientometrics that highlighted the dominance of the United States and the United Kingdom (Sasvári and Lendvai 2025; Cristián et al. 2022), both in terms of publication volume and network centrality. This pattern also highlights a lack of an inclusive scholarly landscape where the field of AI and IP

law remains shaped by Anglo-American legal traditions and institutions. Moreover, the near-total absence of Global South representation stresses enduring inequalities in legal knowledge production (Sasvári and Lendvai 2025).

AI and copyright Law

At the beginning of the period under review, machine learning was being understood and the areas most affected by the technology were being mapped. It soon became clear that the various arts (Chiou 2020), and music in particular (Koempel 2020), were the ones most heavily affected by AI. The medical sector is also strongly influenced and using AI as a writing assistant and the recognizability of the resulting texts is explicitly investigated (Bourré 2023), raising ethical questions as well (Li et al. 2023). The academic literature has explored the possibility and question of whether the development of a particular AI application gives rise to any copyright or related rights, even new ones (Varytimidou 2023), for the AI developer or operator. The question of what the legislator's role is in this new, challenging environment and the effects of slow reaction or over-regulation is also on the academic agenda (Sganga 2023).

The other major – and then defining – strand of copyright issues is the distinction between generative and assisted AI use based on their role in the creative process, and the central question of how to judge the resulting outputs from a copyright point of view (Hugenholtz and Quintais 2021). Fundamental concepts of copyright such as authorship/ownership (Stockton-Brown 2023), originality and the concept of the work itself have been re-examined. The literature on authorship has mainly raised problems in relation to AI-assisted works, asking whether human creativity can be involved in the act of prompting, and where the creative boundaries lie in comparison with ideas (Fenwick & Jurcys 2023): How should EU copyright law handle purely AI-generated works (Trapova 2023)? What will be the impact if the various generated content is put into the public domain without an author (Esayas 2023)? This latter issue also pertained to how distinguishable human creation is to AI-generated content. In this regard, a parallel debate shall also be mentioned on authorship and originality, which is a further strand of the literature that has examined the risk that outputs generated by large language models may themselves infringe existing copyrighted works (see Scharrenberg and Sun 2025). Though this line of research is relatively new and is not present focally in our scientometric analysis, it has to be accentuated since it focuses on one of the most burning questions in the field of IP law and AI, namely problems like substantial similarity, AI training polemics, derivative infringement, and the extent to which generative systems may reproduce protected expression due to memorisation or training data bias (Zhang et al. 2025; Gervais et al. 2024; Liu et al. 2024; Sharma et al. 2025; Al-Busaidi et al. 2024; Nature Editorial 2024, Chesterman 2025; Gaidartzi and Stamatoudi 2025). These concerns are particularly salient in relation to text-, image-, and video-generating models, where stylistic imitation or the reappearance of identifiable fragments raises complex issues of liability allocation between developers, deployers, and users and – as such – the infringement risks associated with AI-generated outputs complement

earlier discussions on the protectability of synthetic content by shifting attention to downstream legal exposure rather than authorship alone.

The input side has also quickly become a prominent concern, in particular the issue of the use of a copyrighted work or a performance protected by neighboring rights when operating AI applications. This implies the examination of possible free uses that differ from one jurisdiction to another, so that the interpretation of the text- and data-mining exception or the fair use test (Xie and He 2023) has proved to be a crucial academic task in this context. The issue of the use of various medical data also arises in the context of opt-out possibilities (McKay et al. 2023). The impact on authors and other related rights holders is also examined, including the issue of remuneration for use (Senftleben 2023). The literature also covers the use of AI as a tool in this process – the identification of works, the possibilities of providing data to rightholders in the context of rights management (Trappey et al. 2020).

AI and digital rights

In the period under review, an increasingly prominent strand of scholarship has examined AI-related challenges through the lens of what is often referred to as “digital rights.” In the context of the present study, digital rights denote legally cognizable interests that arise at the intersection of intellectual property law, personality and publicity rights, data governance, and privacy, particularly where AI systems process, replicate, or simulate human attributes or personal data. In accordance with this, the body of literature presented in this cluster does not stand apart from IP law but instead addresses normative and doctrinal constraints that shape the lawful exploitation, control, and circulation of data, content, and identity in AI-driven environments. Existing research has approached these issues from ethical, legal (Gromova et al. 2023), and social perspectives, while increasingly situating them within broader debates on data-intensive technologies and algorithmic governance (Souza 2018; Chiou 2020; Xie and He 2023).

A central focus within this cluster concerns the use and management of personality-related features (see Chesterman 2020). Such features include the likes of image, voice, facial expressions, and other identifiable human traits induced via AI systems, including deep learning architectures and large language models. Scholarship in this area examines how the replication or simulation of such features implicates personality and publicity-type rights, as well as the boundaries of lawful appropriation in digital environments (Yang and Yu 2022). The concerns outlined in the pertaining literature are strongly connected to IP-adjacent questions, particularly where protected expression, identity, or creative labor is reproduced, stylized, or commercialized through AI-generated outputs.

Closely related debates address the interaction between AI technologies and privacy- and data-protection regimes, which operate as structural limits on the collection, use, and reuse of data for training and deployment purposes. The literature has examined how large-scale data practices intersect with individual rights to privacy and informational self-determination, including compliance challenges arising under data protection frameworks. Specific attention has been paid to emerging phenomena such as “ghostbots” (AI systems trained to simulate the identity, voice, or personality

of deceased or absent individuals) which raise novel questions concerning post-mortem data use, digital identity, and control over personality-linked data in AI-mediated contexts (Harbinja et al. 2023).

Lastly, concerning individual rights, scholarship presented in this cluster has also explored the broader societal implications of AI-driven systems for freedom of expression, access to information, and the diversity of digital environments. Research has highlighted how algorithmic content moderation, recommender systems, and generative models shape visibility and information flows on social media platforms, while simultaneously raising concerns related to misinformation, hallucination, and transparency (Gorwa et al. 2020; Lewis & Moorkens 2020). Concerns also emerge in domains such as academic research, medicine, and education, where AI-assisted practices challenge established norms surrounding authorship, data governance, and the permissible use of protected materials (McKay et al. 2023; Mills et al. 2023).

AI and patent law

Scholarship on artificial intelligence and patent law has also developed along several interconnected lines. In this regard, there are three key issues mentioned in the literature examined.

First, this domain addresses foundational questions of patentability and inventorship in relation to AI-generated inventions. Second, it examines the systemic implications of AI for core patent law doctrines, including novelty and prior art. Third, it explores the use of artificial intelligence as a tool within patent and broader IP systems, both as an aid to administration and as a potential source of new legal risks.

A substantial body of literature focuses on the eligibility of AI-generated inventions for patent protection and the broader impact of AI on the patent system (Rudzite-Celmina 2023; Liberman 2022). From a complementary perspective, scholars have examined whether AI systems themselves could – or should – be designated as inventors, the feasibility of introducing *sui generis* protection for AI-generated inventions (Rudzite-Celmina 2023), and the implications of different inventorship models (Rinkerman 2023). These debates have been shaped significantly by the ongoing global “DABUS saga,” which has resulted in a series of high-profile decisions in jurisdictions including the United States, England and Wales, Germany, Australia, and before the European Patent Office (see Kim 2025). With varying nuances, these decisions have generally rejected the registrability of inventions lacking a human inventor. Parallel scholarship has also raised concerns that AI systems may undermine patentability standards by mass-producing prior art, thereby challenging traditional novelty assessments and prompting proposals to recalibrate existing doctrines (Yordy 2021).

Beyond these foundational issues, research has expanded to examine AI-related patenting activity in emerging technological domains. Studies have addressed patents involving quantum technologies (Kop et al. 2022; Seskir and Willoughby 2023), blockchain and distributed ledger technologies (Denter 2021; Chaudhary 2023), as well as biometric profiling and identification systems (Bakir et al. 2023). Related work has also sought to define and classify key AI technologies, particularly in light

of empirical studies that analyze patent grants and applications submitted to major patent offices (Chang 2020).

Finally, a growing strand of literature considers artificial intelligence as a means of facilitating IP-related processes themselves. Research in this area explores the use of AI tools to support patent examination, classification, and patentability assessment (Lizarralde and Contreras, 2022), as well as analogous applications in trademark law, such as word-embedding approaches to trademark analysis (Shackell and De Vine 2022). Other studies propose deep learning pipelines for automated patent classification (Sofean 2021) or examine whether trained machine learning systems can be conceptualized as protected know-how (Mazzi 2021). At the same time, scholars have raised concerns about transparency, trade secret protection, and potential patent infringement arising from self-learning AI systems, including questions of liability allocation (Mylly 2023; Bohatchuk 2023).

Conclusions, the state of art, and future agenda

To end our paper, the below concluding section synthesizes the findings of the preceding thematic clusters by situating them within recent doctrinal developments in copyright and patent law, before turning to regulatory responses concerning AI training and data use. It then outlines key areas where legal uncertainty persists and where future research is likely to concentrate.

As it was visible from the summary of the previous clusters, the research on the overlaps of AI and IP is not only diverse, but it organically develops in light of the constant technological achievements in the field as well as the changing social needs. While a growing body of judicial and administrative practice emphasizes the requirement of meaningful human involvement for copyright and patent protection, the protectability of AI-generated outputs remains doctrinally contested across jurisdictions and fact patterns (Gervais, 2020; Mezei 2020).

This position is clearly reflected in the numerous judicial and administrative decisions concerning the registrability of inventions generated by the DABUS system developed by Stephen Thaler. In several jurisdictions, including the United States (Thaler v. Vidal, 43 F.4th 1207 (Fed. Cir. 2022)), Australia (Commissioner of Patents v. Thaler [2022] FCAFC 62), England and Wales (Thaler v. Comptroller General of Patents, Trade Marks and Designs [2023] UKSC 49), and before the European Patent Office (Decisions J 8/20 and J 9/20 of the Legal Board of Appeal, 21 December 2021; EPO Examining Division, 25 November 2024), patent protection was refused due to the absence of a human inventor. An important deviation emerged in Germany, where the Federal Supreme Court accepted an amended claim listing both a human applicant and the AI system as inventors (German Federal Supreme Court, 11 June 2024 – Case No. X ZB 5/22). This decision reopened debate on hybrid human–AI inventorship models and has prompted significant scholarly criticism (Kim 2025).

In copyright law, a comparable baseline has been reinforced by U.S. appellate case law and administrative practice, which confirm that at least some degree of human authorship is required in order to benefit from copyright protection.

Scholarly attention, in the copyright discourse more broadly, has also increasingly engaged with issues related to the training phase of AI systems and the lawful use of protected materials as inputs, alongside continuing debates on AI-generated outputs. This is a crucial development which is primarily linked to the emergence of generative AI and its extensive demand for training data, as well as to recent regulatory and policy initiatives. While output-related questions – such as authorship, infringement, and compensation for AI-generated works competing with human creations – remain central, we found that the literature also reflects more and more on the legislative attention in several jurisdictions intensifying around machine learning, data access, and transparency obligations. For instance, the European Union, along with countries such as Japan, Singapore, and the United Kingdom, has taken notable steps toward establishing flexible yet structured frameworks addressing these concerns.

We must remain critical, however, as none of these regulatory approaches are without limitations. Within the European Union, two normative instruments are particularly relevant. First, Directive 2019/790 on copyright and related rights in the Digital Single Market (CDSM Directive) introduced provisions indirectly affecting AI, including text- and data-mining exceptions (Articles 3–4), press publishers' rights (Article 15), and platform liability for user-uploaded content (Article 17). Although these rules are relevant for the AI ecosystem – especially the text- and data-mining exceptions – their scope and practical implications remain contested and normatively uncertain (Ducato and Strowel 2021; Kretschmer, Margoni and Oruç 2024; Mezei 2026), and they may, in some respects, hinder rather than facilitate AI development (Lucchi 2024). Second, the EU's regulatory landscape was complemented by Regulation 2024/1689 (the Artificial Intelligence Act). While not designed as an IP instrument, the AI Act contains copyright-relevant transparency obligations in Article 53, including requirements relating to compliance with text- and data-mining rules and the disclosure of training data summaries. The interaction between these public-law obligations and existing private-law copyright frameworks remains complex and continues to attract scholarly attention (Peukert 2024; Quintais 2025). While these European regulatory developments do feature prominently in contemporary debates, they do not yet constitute a distinct thematic cluster within the existing literature as captured by the present scientometric analysis, mainly because of the novelty of these initiatives, especially the AI Act, thus, substantiate claims of long-term doctrinal consolidation or regulatory stabilization may not be drawn from this analysis alone. Future scientometric research, however, shall undoubtedly reflect on European regulation in much more detail with the body of literature growing exponentially.

While the interface between artificial intelligence and intellectual property law is already governed by existing IP frameworks, the adoption of AI-specific legislative interventions in this area is far from universally accepted. Apart from certain state-level initiatives (such as the Californian Generative Artificial Intelligence: Training Data Transparency Act) and federal-level executive action (most notably the Biden Administration's 2023 Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, which was rescinded shortly after the Second Trump Administration assumed office), the United States has not yet developed a comprehensive or coherent statutory framework specifically addressing the overlap between AI and IP law. On the contrary, current normative approaches remain

fragmented and provide limited *ex ante* guidance for stakeholders. We can interpret this as a restrained regulatory posture which may present a broader reluctance to intervene centrally in the development of AI technologies through binding legislative instruments. Such an approach is reflected in the registration guidelines of the U.S. Patent and Trademark Office (2024) and the U.S. Copyright Office (2023), both of which explicitly recommend maintaining existing registration practices without introducing AI-specific categories. Under these guidelines, the registration of works and inventions remains conditional upon the clear identification of a human author or inventor. This position has been endorsed by judicial practice, including decisions such as *Thaler v. Vidal* (43 F.4th 1207 (Fed. Cir. 2022)) and *Thaler v. Perlmutter* (No. 1:22-cv-01564, D.C. Cir. 2025). Within U.S. intellectual property scholarship, these developments continue to be subject to close scrutiny, generating both supportive arguments (e.g. Brauneis 2025) and sustained criticism (e.g. Lee 2023).

At last, a growing number of pending cases in the United States (at the time of writing well over seventy) concern the use of copyrighted works in the training of generative AI models. Many of these disputes are likely to be dismissed, settled, or otherwise resolved without a definitive ruling on the merits, while others may ultimately be adjudicated by courts or juries. From a scholarly perspective, each outcome contributes to the gradual clarification of doctrinal boundaries rather than representing failure or success in normative terms. At the center of these disputes lies the question of whether the fair use doctrine, a central and highly influential component of U.S. copyright law, can accommodate large-scale machine learning practices. To date, only a few judicial decisions (e.g. *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc.*, D. Del., 2025; *Andrea Bartz, et al. v. Anthropic PBC*,—F. Supp.3d—(2025); *Richard Kadrey, et al. v. Meta Platforms, Inc.* (2025); *The New York Times Co. v. Microsoft Corp.*, et al., 777 F.Supp.3d 283 (2025)) have addressed issues related to AI training and fair use, albeit one of them falls outside the specific context of generative AI. Academic opinion on this matter remains deeply divided. As in earlier debates concerning the protectability of synthetic content, the literature reflects a wide spectrum of views; nevertheless, positions that are cautiously optimistic about the applicability of fair use to machine learning continue to feature prominently (e.g. Lemley and Casey 2021; Murray 2023; Sag 2024; Van Houweling 2025). The institutional landscape has added further complexity to this debate. In May 2025, the U.S. Copyright Office released a pre-publication version of its Generative AI Training report, which, following a detailed and balanced analysis of the fair use doctrine, concluded that “making commercial use of vast troves of copyrighted works to produce expressive content that competes with them in existing markets, especially where this is accomplished through illegal access, goes beyond established fair use boundaries” (United States Copyright Office 2025: 107). One day later, on 10 May 2025, the appointment of the Register of Copyrights, Shira Perlmutter, was terminated, underscoring the politically sensitive and unsettled nature of AI-related copyright governance. (Later, Perlmutter was temporarily reinstated to her position; while the U.S. Supreme Court, in an unsigned order, indicated that it would not pause the ruling of the federal appeals court decision on the reinstatement.) These developments suggest that forthcoming policy initiatives, including the anticipated U.S.

President's AI Action Plan, may yet recalibrate the balance between innovation and rights protection.

Limitations

Although this study offers a comprehensive scientometric overview of the scholarly discourse on artificial intelligence and intellectual property law, several limitations must be acknowledged. In this sense, the present analysis maps structured visibility within Scopus-indexed journal literature rather than the full trajectory of doctrinal innovation, which in legal scholarship often emerges in working papers, books, and specialized law reviews prior to indexed journal publication.

First and foremost, the analysis was based exclusively on data retrieved from the Scopus database. Although Scopus is widely recognized for its extensive coverage of peer-reviewed literature, particularly in the social sciences, it does not equally represent all legal scholarship (Sasvári and Lendvai 2025). Legal research is frequently disseminated through specialized journals, books, edited volumes, scientific and law-related blogs and non-academic platforms, as well as non-indexed national law reviews, many of which may not be included in Scopus. As such, this study does not claim to capture the entirety of global legal scholarship on AI and IP law, but rather maps patterns of thematic development, collaboration, and visibility within Scopus-indexed, English-language journal publications, where citation-based scientometric analysis is methodologically and scientometrically feasible. The above limitation has particular implications for the geographic distribution of authorship observed in the study. Legal scholarship originating from the Global South or from non-English-speaking jurisdictions is often published in locally oriented outlets, which are less likely to be indexed in large international citation databases. Consequently, the under-representation of Global South-affiliated authors identified in the co-authorship analysis should be interpreted as reflecting disparities in visibility and inclusion within the specific bibliometric corpus (in our case, Scopus) examined, rather than the absence of relevant scholarship “*per se*.” Accordingly, both the thematic clusters identified in this study and the distribution of scholars regionally reflect patterns of visibility and concentration within the indexed, specific corpus we used, thereby, they are not definitive measures of doctrinal influence or innovation across the broader field of AI and IP law. Second, though the methodological decision to restrict the search to final-stage (i.e., “actually” published) journal articles published between 2013 and 2023 ensured a high degree of comparability and metadata reliability, it may also have excluded relevant scholarship, such as preprints on different yet well-known scholarly platforms, legal opinions on well-renowned legal blogs, legal conference proceedings, or policy documents indexed perhaps in other lesser used data repositories in scientometric research that contribute meaningfully to the evolving debate on AI and IP law. We acknowledge this limitation as legal scholarship frequently circulates through preprint repositories and working paper series prior to (or instead of) formal journal publication (see Sasvári and Lendvai, 2025). Future research could therefore extend the present analysis by incorporating preprints hosted on platforms such as SSRN or other repositories that serve as prominent venues for

legal scholarship, thereby providing a more comprehensive mapping of the field. Third, while the thematic analysis based on keyword co-occurrence and clustering does show configurations of topical proximity and structural prominence within the indexed corpus, it does not independently differentiate between instrumental doctrinal interventions and shorter, trend-responsive contributions. In this sense, network measures highlight connectivity and visibility within the network which are not interchangeable with intellectual leadership or long-term canonical influence. Thus, the results should be interpreted as mapping thematic organization and adjacency, not as establishing doctrinal hierarchy.

Appendix 1

<https://docs.google.com/document/d/13wT6uzgZXbg4hBUqytsvdVHN9iVNUf4e/e/dit?usp=sharing&oid=109536681888630160892&rtpof=true&sd=true>

Author contributions All authors made substantial contribution to the concept of the work, analysis, interpretation of data, use of software, data collection and extraction. All authors have drafted the work and approved the version to be published. All authors agree to be accountable for all aspects of the work.

Funding Open access funding provided by Ludovika University of Public Service. Project TKP2021-NKTA-51 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NKTA funding scheme. The research was made with the financial support of a fellowship from the Center for Advanced Internet Studies (CAIS).

Data availability Data is made available upon request. Additional data can be accessed under Appendix 1.

Declarations

Ethics approval Not applicable

Permission to reproduce material from other sources Not applicable

Conflict of interest The authors declare no conflict of interest

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Al-Busaidi AS, Raman R, Hughes L, Albashrawi MA, Malik T, Dwivedi YK, Al-Alawi T, AlRizeiqi M, Davies G, Fenwick M, Gupta P, Gurpur S, Hooda A, Jurcys P, Lim D, Lucchi N, Misra T, Raman R, Shirish A, Walton P (2024) Redefining boundaries in innovation and knowledge domains: investigating the impact of generative artificial intelligence on copyright and intellectual property rights. *J Educ Chang Innov Knowl* 9(4):100630. <https://doi.org/10.1016/j.jik.2024.100630>
- Baas J, Schotten M, Plume A et al (2020) Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quant Sci Stud* 1:377–386. https://doi.org/10.1162/qss_a_00019
- Bakir V, Laffer A, McStay A (2023) Human-First, please: Assessing citizen views and industrial ambition for emotional AI in recommender systems. *Surveillance & Society* 21:205–222. <https://doi.org/10.24908/ss.v21i2.16015>
- Björk B-C, Solomon D (2013) The publishing delay in scholarly peer-reviewed journals. *J Informetrics* 7(4):914–923. <https://doi.org/10.1016/j.joi.2013.09.001>
- Bohatchuk D (2023) Transparency as a legal value for patent disclosure. *JIPITEC* 14:190–210
- Bonadio E, McDonagh L (2020) Artificial intelligence as producer and consumer of copyright works: evaluating the consequences of algorithmic creativity. *SSRN Electronic Journal*
- Bourré N (2023) Can readers spot the AI impostor in healthcare writing? *Med Writ* 32(3):38–40. <https://doi.org/10.56012/fwhk6920>
- Brauneis R (2025) Copyright and the training of human authors and generative machines. *JLA* 48(1). <https://doi.org/10.52214/jla.v48i1.13529>
- Calboli I (2021) Comparative legal analysis and intellectual property law: a guide for research. Oxford University Press eBooks, pp 46–66
- Cataldo I, Lieu AA, Carollo A, Bornstein MH, Gabrieli G, Lee A, Esposito G (2022) From the cradle to the Web: the growth of “sharenting”—A scientometric perspective. *Hum Behav Emerg Technol* 2022:1–12. <https://doi.org/10.1155/2022/5607422>
- Chang S-H (2020) Technical trends of artificial intelligence in standard-essential patents. *Data Technologies and Applications* 55:97–117. <https://doi.org/10.1108/dta-10-2019-0178>
- Chaudhary N (2023) Artificial intelligence and Blockchain: a breakthrough collaboration in IP law. *J Retailing Intellect Property Rights* 28. <https://doi.org/10.56042/jipr.v28i5.981>
- Chen C (2006) CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. *J Am Soc Inf Sci* 57(3):359–377. <https://doi.org/10.1002/asi.20317>
- Chesterman S (2020) Artificial intelligence and the limits of legal personality. *ICLQ* 69(4):819–844. <https://doi.org/10.1017/s0020589320000366>
- Chesterman S (2025) Good models borrow, great models steal: intellectual property rights and generative AI. *Policy Soc* 44(1):23–37. <https://doi.org/10.1093/polsoc/puae006>
- Chiou T (2020) Copyright lessons on machine learning: what impact on algorithmic art? *JIPITEC* 10:398–411
- Christián L, Háló G, Demeter M (2022) Twenty years of law journal publishing: a Comparative analysis of international publication trends. *Pub Res Q* 38(1):1–17. <https://doi.org/10.1007/s12109-021-09854-1>
- Davies CR (2011) An evolutionary step in intellectual property rights – artificial intelligence and intellectual property. *Comput Law Secur Rev* 27(6):601–619. <https://doi.org/10.1016/j.clsr.2011.09.006>
- Demeter M (2020) Academic knowledge production and the Global South. Palgrave Macmillan. Cham.
- Denter NM (2021) Blockchain breeding grounds: Asia’s advance over the USA and Europe. *World Patent Information* 67:102082. <https://doi.org/10.1016/j.wpi.2021.102082>
- Ding X, Yang Z (2022) Knowledge mapping of platform research: a visual analysis using VOSviewer and CiteSpace. *Electron Commer Res* 22(3):787–809. <https://doi.org/10.1007/s10660-020-09410-7>
- Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM (2021) How to conduct a bibliometric analysis: an overview and guidelines. *J Bus Res* 133:285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Ducato R, Strowel A (2021) Ensuring text and data mining: remaining issues with the EU copyright exceptions and possible ways out. *European Intellectual Property Review* 43:322–337
- Elkin-Koren N (2020) Contesting algorithms: restoring the public interest in content filtering by artificial intelligence. *Big Data Soc* 7(2). <https://doi.org/10.1177/2053951720932296>
- Erb RA (2022) Does artificial intelligence (AI) have a role in E-Resources licensing? *Ser Libr* 82(1–4):83–90. <https://doi.org/10.1080/0361526x.2022.2018240>

- Esayas S (2023) The important role of emergence in conceptualizing the challenges of new technologies to private law. *European Review of Private Law/Revue Européenne De Droit Privé/Europäische Zeitschrift Für Privatrecht* 31:779–822. <https://doi.org/10.54648/erpl2023034>
- Fenwick M, Jurecs P (2023) Originality and the future of copyright in an age of generative AI. *Computer Law & Security Review* 51:105892. <https://doi.org/10.1016/j.clsr.2023.105892>
- Gaidartzi A, Stamatoudi I (2025) Authorship and ownership issues raised by AI-Generated works: a Comparative analysis. *Laws* 14(4):57. <https://doi.org/10.3390/laws14040057>
- Gervais DJ (2020) The machine as author. *Iowa Law Rev* 105:2053–2106
- Gervais DJ, Shemtov N, Marmanis H, Rowland CZ (2024) The heart of the matter: copyright, AI training, and LLMs. *SSRN Electron J*. <https://doi.org/10.2139/ssrn.4963711>
- Glasser D (2001) Copyrights in computer-generated works: whom, if anyone, do we reward? *Duke law scholarship repository (Duke University)* 1:24
- Gorwa R, Binns R, Katzenbach C (2020) Algorithmic content moderation: technical and political challenges in the automation of platform governance. *Big Data Soc* 7(1):205395171989794. <https://doi.org/10.1177/2053951719897945>
- Grant MJ, Booth A (2009) A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libraries J* 26(2):91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Gromova EA, Ferreira DB, Begishev IR (2023) ChatGPT and other intelligent chatbots: legal, ethical and dispute resolution concerns. *Revista Brasileira de Alternative Dispute Resolut* 5(10). <https://doi.org/10.52028/rbadr.v5i10.art07.ru>
- Gupta P, Wadhwa S, Chauhan S (2023) Crossroad of intellectual property rights of technology innovators and human rights: a systematic literature review. *DPRG* 25(3):236–249. <https://doi.org/10.1108/dprg-08-2022-0099>
- Harbinja E, Edwards L, McVey M (2023) Governing ghostbots. *Comput Law Secur Rev* 48:105791. <https://doi.org/10.1016/j.clsr.2023.105791>
- Hart RJ (1988) Copyright and computer generated works. In *Aslib Proceedings*, vol 40. pp 173–181. <https://doi.org/10.1108/eb051098>
- Higham KW, Governale M, Jaffe AB, Zülicke U (2017) Unraveling the dynamics of growth, aging and inflation for citations to scientific articles from specific research fields. *J Informetrics* 11(4):1190–1200. <https://doi.org/10.1016/j.joi.2017.10.004>
- Hugenholtz PB, Quintais JP (2021) Copyright and artificial creation: does EU copyright law protect AI-Assisted output? *IIC* 52(9):1190–1216. <https://doi.org/10.1007/s40319-021-01115-0>
- Hunter JD (2007) MatPlotLib: a 2D graphics environment. *Comput Sci Eng* 9(3):90–95. <https://doi.org/10.1109/mcse.2007.55>
- Kahveci ZÜ (2023) Attribution problem of generative AI: a view from US copyright law. *J Intellect Property Law Pract* 18(11):796–807. <https://doi.org/10.1093/jiplp/jpad076>
- Kalpokiene J, Kalpokas I (2023) Creative encounters of a posthuman kind – anthropocentric law, artificial intelligence, and art. *Technol Soc* 72:102197. <https://doi.org/10.1016/j.techsoc.2023.102197>
- Khan SN, Loukil F, Ghedira-Guegan C, Benkhelifa E, Bani-Hani A (2021) Blockchain smart contracts: applications, challenges, and future trends. *Peer-To-Peer Netw Appl* 14(5):2901–2925. <https://doi.org/10.1007/s12083-021-01127-0>
- Kim D (2025) The illusory standard of significant human contribution to AI-Assisted inventions after the DABUS decision of the German Federal Court of justice. *IIC* 56(2):369–380. <https://doi.org/10.1007/s40319-025-01567-8>
- Koempel F (2020) From the gut? Questions on artificial intelligence and music. *QMJP* 10(4):503–513. <https://doi.org/10.4337/qmjp.2020.04.05>
- Kop M, Abov M, Minssen T (2022) Patenting Quantum Technologies & Market Power: A Quantitative Analysis, Conference Paper, The Legal Dimensions of Quantum Computing, Lund University
- Kretschmer M, Margoni T, Oruç P (2024) Copyright law and the lifecycle of machine learning models. *IIC* 55:110–138. <https://doi.org/10.1007/s40319-023-01419-3>
- Lee E (2023) Prompting progress: authorship in the age of AI. *Ssrn J*. <https://doi.org/10.2139/ssrn.4609687>
- Lemley MA, Casey B (2021) Fair learning. *Ssrn J* 99:743–785. <https://doi.org/10.2139/ssrn.3528447>
- Lendvai G (2025) Growth and visibility of Latin American and Caribbean law journals (2008–2023): a focus on their language, collaboration, and open access trends. *Iberoamerican J Sci Meas Commun* 5(1):1–20. <https://doi.org/10.47909/ijsmc.149>
- Levine AM, Heckman JD, Hensinger RN (2004) The art and science of reviewing manuscripts for orthopaedic journals: part I. Defining the review. *Instr Course Lect* 53:679–688

- Lewis D, Moorkens J (2020) A Rights-Based approach to trustworthy AI in social media. *Social Media + Society* 6. <https://doi.org/10.1177/2056305120954672>
- Li J, Cai X, Cheng L (2023) Legal regulation of generative AI: a multidimensional construction. *Int J Multiling Legal Discourse* 8(2):365–388. <https://doi.org/10.1515/ijld-2023-2017>
- Liberman A (2022) One small step for ‘artificial intelligence’ and a giant leap for the Australian patent system? The Federal Court decision in *Thaler v Commissioner of patents*. *J Intellectual Property Law Pract* 17(2):164–178. <https://doi.org/10.1093/jiplp/jpab181>
- Liu X, Sun T, Xu T et al (2024) SHIELD: evaluation and defense strategies for copyright compliance in LLM text generation. arXiv 1640-1670. <https://doi.org/10.18653/v1/2024.emnlp-main.98>
- Lizarralde MD, Contreras HA (2022) The real role of AI in patent law debates. *International Journal of Law and Information Technology* 30:23–46. <https://doi.org/10.1093/ijlit/eaac008>
- Lone AH, Naaz R (2021) Applicability of Blockchain smart contracts in securing Internet and IoT: a systematic literature review. *Comput Sci Rev* 39:100360. <https://doi.org/10.1016/j.cosrev.2020.100360>
- Lucchi N (2024) ChatGPT: a case study on copyright challenges for generative artificial intelligence systems. *Eur J Risk Regul* 15(3):602–624. <https://doi.org/10.1017/err.2023.59>
- Mammen CE, Richey C (2020) AI and IP: are creativity and inventorship inherently human activities? *Lawreview* 14(2). <https://doi.org/10.25148/lawrev.14.2.9>
- Manterola C, Astudillo P, Arias E, Claros N (2013) Systematic reviews of the literature: what should be known about them. *Cirugía Española (English Edition) Cirugía Española (Engl Ed)* 91(3):149–155. <https://doi.org/10.1016/j.cireng.2013.07.003>
- Mazzi F (2021) CMOs and AI: Can trained machine learning be justified with the concept of know-how? *World Patent Information* 65:102036. <https://doi.org/10.1016/j.wpi.2021.102036>
- McKay F, Treanor D, Hallowell N (2023) Inalienable data: ethical imaginaries of de-identified health data ownership. *SSM - Qualitative Res Health* 4:100321. <https://doi.org/10.1016/j.ssmqr.2023.100321>
- Meho LI (2007) The rise and rise of citation analysis. *Phys World* 20(1):32–36. <https://doi.org/10.1088/2058-7058/20/1/33>
- Mezei P (2020) From Leonardo to the next rembrandt – the need for AI-Pessimism in the age of Algorithms. *UFITA - Arch Für Medienrecht Und Medienwissenschaft* 84(2):390–429. <https://doi.org/10.5771/2568-9185-2020-2-390>
- Mezei P (2026) Five years after: A critical re-evaluation of end-user flexibilities under the CDSM Directive. In: Sganga, Caterina; Synodinou, Tatiana Eleni (eds.) *Flexibilities in Copyright Law*. pp 63–80
- Miller AR (1993) Copyright protection for computer programs, databases, and computer-generated works: is anything new since CONTU? *Harv Law Rev* 106(5):977. <https://doi.org/10.2307/1341682>
- Mills A, Bali M, Eaton L (2023) How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process. *J Appl Learn Teach* 6. <https://doi.org/10.37074/jalt.2023.6.1.34>
- Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E (2018) Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* 18(1). <https://doi.org/10.1186/s12874-018-0611-x>
- Murray MD (2023) Generative AI art: copyright infringement and fair use. *STLR* 26(2):259. <https://doi.org/10.25172/smustr.26.2.4>
- Mylly U-M (2023) Transparent AI? Navigating between rules on trade secrets and access to information. *IIC* 54(7):1013–1043. <https://doi.org/10.1007/s40319-023-01328-5>
- Nature Editorial (2024) AI firms must play fair when they use academic data in training. *Nature* 632(8027):953. <https://doi.org/10.1038/d41586-024-02757-z>
- Norris M, Oppenheim C (2007) Comparing alternatives to the Web of Science for coverage of the social sciences’ literature. *J Informetrics* 1(2):161–169. <https://doi.org/10.1016/j.joi.2006.12.001>
- Offenhuber D (2024) Shapes and frictions of synthetic data. *Big Data Soc* 11(2). <https://doi.org/10.1177/20539517241249390>
- Peukert A (2024) Copyright in the artificial intelligence Act – a primer. *GRUR Int* 73(6):497–509. <https://doi.org/10.1093/grurint/ikae057>
- Picht PG, Thouvenin F (2023) AI and IP: theory to policy and back again – policy and research recommendations at the intersection of artificial intelligence and intellectual property. *IIC* 54(6):916–940. <https://doi.org/10.1007/s40319-023-01344-5>
- Qiao T, Ma Y, Zheng N, Wu H, Chen Y, Xu M, Luo X (2023) A novel model watermarking for protecting generative adversarial network. *Comput Secur* 127:103102. <https://doi.org/10.1016/j.cose.2023.103102>

- Quintais JP (2025) Generative AI, copyright and the AI Act. *Comput Law Secur Rev* 56:106107. <https://doi.org/10.1016/j.clsr.2025.106107>
- Rinkerman G (2023) Artificial Intelligence and evolving issues under US copyright and patent law. *Interactive Entertainment Law Review* 6:48–65. <https://doi.org/10.4337/ielr.2023.0002>
- Rudzite-Celmina L (2023) Patent-eligible invention requirement under the European patent convention and its implications on creations involving artificial intelligence. *MUJLT* 17(2):249–279. <https://doi.org/10.5817/mujlt2023-2-4>
- Sag M (2024) Fairness and fair use in generative AI. *Fordham Law Rev* 92:1887–1921
- Samuelson P (1986) Allocating ownership rights in computer-generated works. *Univ Pittsbg Law Rev* 47:1185–1228
- Sasvári P, Lendvai GF (2025) The overrepresentation of the United States in the field of legal studies in the science-wide author databases of standardized citation indicators. *J Informetrics* 19(3):101680. <https://doi.org/10.1016/j.joi.2025.101680>
- Schafer B, Komuves D, Zatarain JMN, Diver L (2015) A fourth law of robotics? Copyright and the law and ethics of machine co-production. *Artif Intell Law* 23(3):217–240. <https://doi.org/10.1007/s10506-015-9169-7>
- Scharrenberg N, Sun C (2025) Copyright infringement by large language models in the EU: misalignment, safeguards, and the path forward. In *Proceedings of the Natural Legal Language Processing Workshop*, pp 125–134. <https://doi.org/10.18653/v1/2025.nllp-1.9>
- Senftleben M (2023) Generative AI and author remuneration. *IIC* 54(10):1535–1560. <https://doi.org/10.1007/s40319-023-01399-4>
- Seskir ZC, Willoughby KW (2023) Global innovation and competition in quantum technology, viewed through the lens of patents and artificial intelligence. *International Journal of Intellectual Property Management* 13:40. <https://doi.org/10.1504/ijipm.2023.129076>
- Sganga C (2023) Is there still a policy agenda for EU copyright law? *IIC* -. *IIC* 54(9):1407–1417. <https://doi.org/10.1007/s40319-023-01365-0>
- Shackell C, De Vine L (2022) Quantifying the genericness of trademarks using natural language processing: an introduction with suggested metrics. *Artif Intell Law* 30(2):199–220. <https://doi.org/10.1007/s10506-021-09291-7>
- Sharma AS, Sanyal D, Srivastava P et al (2025) Nine ways to break copyright law and why our LLM won't: a fair use aligned generation framework. *Arxiv* (Cornell University). <https://doi.org/10.48550/arxiv.2505.23788>
- Skvortsova TA, Nesterova NV, Shatveryan NG et al (2023) Artificial intelligence as an object of civil law regulation. In: *Smart innovation, systems and technologies* 127–135
- Sofean M (2021) Deep learning based pipeline with multichannel inputs for patent classification. *World Patent Information* 66:102060. <https://doi.org/10.1016/j.wpi.2021.102060>
- Souza RR (2018) Algorithms, future and digital rights: some reflections. *EFI* 34(3):179–183. <https://doi.org/10.3233/efi-180200>
- Stockton-Brown M (2023) Inking cultures: authorship, AI-Generated art and copyright law in tattooing. *Int J Semiot Law* 36(5):2037–2065. <https://doi.org/10.1007/s11196-023-10024-z>
- Tătaru G-C, Domenteanu A, Delcea C, Florescu MS, Orzan M, Cotfas L-A (2024) Navigating the disinformation maze: a bibliometric analysis of scholarly efforts. *Information* 15(12):742. <https://doi.org/10.3390/info15120742>
- Trapova A (2023) Copyright for AI-generated works: a task for the internal market? *European Law Review* 48:187–205
- Trappey AJC, Lupu M, Stjepandic J (2020) Embrace artificial intelligence technologies for advanced analytics and management of intellectual properties. *World Patent Inf* 61:101970. <https://doi.org/10.1016/j.wpi.2020.101970>
- Van Eck NJ, Waltman L (2017) Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics* 111(2):1053–1070. <https://doi.org/10.1007/s11192-017-2300-7>
- U.S. Copyright Office (2023) Copyright registration guidance: works containing material generated by artificial intelligence.
- U.S. Copyright Office (2025) Copyright and artificial intelligence Part, 3: generative AI training
- U.S. Patent and Trademark Office (2024) 2024 Guidance update on patent subject matter eligibility, including on artificial intelligence (2024). <https://www.federalregister.gov/documents/2024/07/17/2024-15377/2024-guidance-update-on-patent-subject-matter-eligibility-including-on-artificial-intelligence>. in: *Federal Register*
- Van Houweling MS (2025) The freedom to extract in copyright law. *N Carol Law Rev* 103:445–520

- Varytimidou C (2023) The New A(I)rt Movement and Its Copyright Protection: Immoral or E-Moral? *GRUR International* 72:357–363. <https://doi.org/10.1093/grurint/ikac153>
- Wach K, Duong CD, Ejdyś J, Kazlauskaitė R, Korzynski P, Mazurek G, Paliszkievicz J, Ziemia E (2023) The dark side of generative artificial intelligence: a critical analysis of controversies and risks of ChatGPT. *EBER* 11(2):7–30. <https://doi.org/10.15678/eber.2023.110201>
- Weber-Wulff D, Anohina-Naumeca A, Bjelobaba S, Foltýnek T, Guerrero-Dib J, Popoola O, Šigut P, Waddington L (2023) Testing of detection tools for AI-generated text. *Int J Educ Integr* 19(1). <https://doi.org/10.1007/s40979-023-00146-z>
- Xie Q, He T (2023) Battle between ‘long’ and ‘Short’ videos: fragmented uses, diversified purposes, and the evolution of China’s copyright limitation rules in the AI Era. *The Chin J Comp Law* 11(3). <https://doi.org/10.1093/cjcl/cxae001>
- Yang Y, Yu D (2022) Short video copyright storage algorithm based on blockchain and expression recognition. *Int J Soc Res Methodol Digit Multimedia Broadcast* 2022:1–11. <https://doi.org/10.1155/2022/8827815>
- Yordy LR (2021) The Library of Babel for Prior Art: Using Artificial Intelligence to Mass produce Prior Art in Patent Law. *Vanderbilt Law Review* 74:521
- Zhang D, Xu Z, Zhao W (2025) Llms and copyright risks: benchmarks and mitigation approaches. In *Proceedings of the 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies: Tutorial Abstracts*, pp 44–50. <https://doi.org/10.18653/v1/2025.naacl-tutorial.7>
- Zhu J, Liu W (2020) A tale of two databases: the use of Web of Science and Scopus in academic papers. *Scientometrics* 123(1):321–335. <https://doi.org/10.1007/s11192-020-03387-8>
- Zyhrii O, Trufanova Y, Parashchuk L, Sampara N, Tsvigun I (2023) Law and technology: the impact of innovations on the legal system and its regulation. *SLS* 6(4):267–275. <https://doi.org/10.32518/sals.4.2023.267>