Page: 347-382 *URL: www.lijdlr.com*

LAWFOYER INTERNATIONAL JOURNAL OF DOCTRINAL LEGAL RESEARCH

ISSN: 2583-7753

LawFoyer International Journal of Doctrinal Legal Research

[ISSN: 2583-7753]

Volume IV | Special Issue I

2026

Special Issue on the International E-Conference on

"Law in the Age of Artificial Intelligence: A Fundamental Perspective"

Organised by KES' Shri J. H. Patel College of Law, Mumbai, on 24 January 2026

Official Publication Partner: LawFoyer International Journal of Doctrinal Legal Research

DOI: https://doi.org/10.70183/lijdlr.2026.v4si1.21

KES' Shri. Jayantilal H. Patel Law College

About the College

KES' Shri. Jayantilal H. Patel Law College is committed to nurturing and empowering students through quality legal education. The college provides an inclusive and academically enriching environment that encourages students to excel in their studies, participate in co-curricular activities, engage in research, and develop a strong sense of civic responsibility.

The college is affiliated to the University of Mumbai, approved by the Bar Council of India, NAAC accredited with 'B++' grade, and ISO 9001:2015 certified. It offers programmes including B.A. LL.B., LL.B., LL.M., and specialised diploma courses.

With qualified and experienced faculty, a strong academic environment, moot court training, research activities and modern infrastructure, the institution seeks to prepare responsible, skilled and socially conscious legal professionals equipped to meet the challenges of a rapidly changing world.

The college functions under the aegis of Kandivli Education Society (KES), an institution established in 1936 with a long-standing commitment to quality, inclusivity and value-based education.

Affiliation	University of Mumbai
Recognition	Approved by the Bar Council of India
Accreditation	NAAC accredited 'B++'
Certification	ISO 9001:2015 certified

Kandivli Education Society

Established 1936

Kandivli Education Society (KES), established in 1936, has grown from a small educational initiative into a dynamic educational institution serving thousands of students.

With a legacy of quality, inclusivity and value-based learning, KES has expanded its academic presence across various disciplines. Its educational institutions are committed to providing meaningful learning opportunities and fostering academic, professional and personal development among students.

KES' Shri. Jayantilal H. Patel Law College is one of its prominent institutions dedicated to legal education. Through academic excellence, research, practical learning and co-curricular engagement, the college seeks to empower students with the knowledge, skills and values necessary to contribute meaningfully to society.

Educational Commitment

- Quality and inclusive education
- Value-based learning
- Academic and professional development
- Research and practical learning
- Holistic and future-ready education

Law in the Age of Artificial Intelligence

A Fundamental Perspective

About the International Conference

The International Conference on "Law in the Age of Artificial Intelligence: A Fundamental Perspective" provided an academic platform to examine the transformative impact of Artificial Intelligence on law, legal systems, governance and society.

The conference explored emerging legal and ethical questions relating to Artificial Intelligence, including intellectual property, authorship and ownership, liability, data privacy, cybersecurity, criminal justice, evidence law, human rights, ethics and accountability. It addressed the implications of Artificial Intelligence for legal education, commercial law, regulatory frameworks, policy development, research and scholarly communication.

The conference brought together academicians, research scholars, legal professionals and students to present research, exchange ideas and engage in interdisciplinary discussions. It encouraged meaningful academic dialogue and contributed towards the development of effective, ethical and inclusive legal responses to the rapidly evolving Artificial Intelligence landscape.

Conference	International Conference
Theme	Law in the Age of Artificial Intelligence: A Fundamental Perspective
Date	24 January 2026
Time	10:00 AM onwards
Mode	Online
Organised by	IQAC, Student Research Society and Library Committee
Publication Partner	LawFoyer International Journal of Doctrinal Legal Research [ISSN: 2583-7753]

Principal's Message

KES' Shri. Jayantilal H. Patel Law College

It gives me immense pleasure to welcome all academicians, research scholars, legal professionals and students to the International Conference on "Law in the Age of Artificial Intelligence: A Fundamental Perspective."

Artificial Intelligence is transforming the way societies function and is increasingly influencing legal systems, governance, professional practices and education. These developments bring significant opportunities as well as complex legal, ethical and social challenges.

This conference provided a valuable platform for intellectual exchange and interdisciplinary dialogue on emerging issues such as data privacy, intellectual property, liability, criminal justice, human rights, ethics and the governance of Artificial Intelligence. The scholarly contributions and discussions during the conference encouraged deeper understanding and promoted responsible approaches to the use of Artificial Intelligence.

I extend my best wishes to all the speakers, researchers, presenters, participants and members of the organising team for a meaningful and intellectually enriching conference.

Dr. Viral Dave
I/c. Principal
KES' Shri. Jayantilal H. Patel Law College

Organising Committee

International Conference on Law in the Age of Artificial Intelligence

I/c. Principal	Dr. Viral Dave
Convenor	Mr. Mahavir Mandot
Co-Convenors	Dr. Shweta Vijendra Pathak Ms. Priyanka Khakhadia Ms. Deepanshi Chandarana

Organised by IQAC, Student Research Society and Library Committee

KES' Shri. Jayantilal H. Patel Law College, Mumbai

Papers presented at the conference are published in the Special Issue of the LawFoyer International Journal of Doctrinal Legal Research [Volume IV, Special Issue I, 2026]

COPYRIGHT CAPTURE IN THE AI AGE: BIG TECH, ACADEMIC PAYWALLS, AND THE TRIPS DILEMMA

Mahavir Mandot[1] & Dr. Renupal Sood[2]

I. ABSTRACT

The rapid advancement of artificial intelligence (AI), particularly large language models and generative AI systems, has transformed the production, dissemination, and commercial use of knowledge in the digital era. These systems rely on vast datasets drawn from books, research articles, websites, and other digital materials, many of which may be protected by copyright. While such datasets have enabled technological innovation, they have also raised serious legal and ethical concerns regarding authorization, compensation, ownership of training data, and the commercialization of intellectual works. This paper examines the phenomenon of "copyright capture," whereby large technology corporations derive economic value from copyrighted materials through AI training without clearly defined licensing or benefit-sharing mechanisms. The issue is intensified by the dominance of companies possessing the infrastructure, computational capacity, and financial resources required to develop large-scale AI models. At the same time, academic publishers continue to restrict access to scholarly literature through subscription-based paywalls, creating a paradox in which AI systems may process knowledge that remains inaccessible to many human researchers, particularly in developing countries and under-resourced academic institutions. Using a doctrinal legal research methodology, the paper analyses copyright principles, academic publishing practices, the TRIPS Agreement, and emerging AI governance developments. It argues that existing international copyright regimes inadequately address AI training datasets, automated data extraction, and AI-generated outputs. The paper concludes that transparent data governance, fair licensing frameworks, stronger open access

[1] Research Scholar, Faculty of Legal Studies, Shoolini University, Solan, Himachal Pradesh (India).Email: mahavirmandot97@gmail.com
[2] LLM (Business Law), Faculty of Legal Sciences, Shoolini University, Solan (India).Email: mahavirmandot97@gmail.com

 (ISSN: 2583-7753)

policies, and coordinated international cooperation are necessary to balance technological innovation with creators' rights and equitable access to knowledge in the AI age.

II. KEYWORDS

Artificial Intelligence, Copyright Capture, Academic Publishing, TRIPS Agreement, Data Governance.

III. INTRODUCTION

A. Rise of Artificial Intelligence Technologies

Artificial Intelligence (AI) has emerged as one of the most transformative technological developments of the twenty-first century. Over the past decade, rapid advancements in machine learning, neural networks, and computational infrastructure have enabled the creation of sophisticated AI systems capable of performing complex cognitive tasks such as language generation, image recognition, and predictive analysis. These technologies are now widely used across diverse sectors including healthcare, finance, education, law, and governance.

[3]One of the most significant developments in this field has been the emergence of generative AI models capable of producing human-like text, images, and audio outputs. These systems rely on large-scale machine learning architectures that process vast quantities of digital information to identify patterns and generate new content. As a result, access to massive datasets has become a critical factor in the development and success of modern AI technologies.[4]

B. Importance of Training Datasets in AI Systems

AI systems do not develop intelligence independently; rather, they rely on extensive training datasets that allow algorithms to learn patterns, linguistic structures, and

[3] Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
[4] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press

contextual relationships. These datasets often include digital content sourced from books, academic journals, websites, and social media platforms.[5]

The process of collecting and processing such large datasets frequently involves automated techniques such as web scraping and data mining. While these practices have accelerated technological innovation, they also raise important legal questions regarding the use of copyrighted materials for machine learning purposes. Many datasets used for training AI models contain copyrighted works whose creators may not have provided explicit authorization for such usage.

C. Increasing Use of Copyrighted Content in AI Training

The integration of copyrighted materials into AI training datasets has triggered significant debates within the fields of intellectual property law and digital governance. AI developers often argue that the use of publicly available digital content constitutes legitimate data analysis or falls under copyright exceptions such as fair use or text-and-data mining. However, copyright holders frequently contend that such large-scale data extraction constitutes unauthorized reproduction of protected works.[6]

These debates have become increasingly prominent as generative AI platforms demonstrate the ability to produce outputs that closely resemble existing copyrighted works. This raises concerns regarding authorship, ownership, and the potential displacement of human creators within creative industries.

D. Emergence of the Concept of Copyright Capture

The concept of "copyright capture" has emerged in recent academic discussions to describe situations where large technology companies derive economic value from copyrighted content without establishing equitable compensation mechanisms for creators. By training AI models on massive collections of copyrighted materials,

[5] Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux.
[6] Samuelson, P. (2023). Generative AI meets copyright. *Columbia Journal of Law & the Arts*, 46(2)

technology firms can develop highly profitable products while the original authors receive little or no financial recognition.[7]

This phenomenon is closely linked to the concentration of technological power within a small number of global technology corporations that possess the computational resources and financial capital required to develop large-scale AI systems.

E. Research Objectives

The present research is guided by the following objectives:

1. To examine the legal implications of using copyrighted works in artificial intelligence training datasets, particularly in relation to reproduction rights, fair use, and text-and-data mining exceptions.
2. To evaluate the adequacy of the TRIPS Agreement, especially Article 13 and the three-step test, in addressing copyright issues arising from AI training practices and automated data extraction.
3. To analyse the relationship between academic publishing paywalls, the AI knowledge paradox, and unequal access to scholarly knowledge in developing countries.
4. To propose transparent data governance, fair licensing, and open access reforms that balance technological innovation with the protection of authors' and researchers' rights.

F. Research Questions

The study seeks to answer the following research questions:

1. Whether the use of copyrighted materials in artificial intelligence training datasets can be justified under existing copyright exceptions such as fair use or text-and-data mining?

[7] Lemley, M. A., & Casey, B. (2021). Fair learning. *Texas Law Review*, 99.

2. Whether Article 13 of the TRIPS Agreement and the three-step test provide sufficient flexibility to address AI training, automated data extraction, and AI-generated outputs?
3. How do academic paywalls contribute to the AI knowledge paradox, particularly where artificial intelligence systems may process scholarly materials that remain inaccessible to many human researchers?
4. What licensing, transparency, and open access reforms are necessary to prevent copyright capture while promoting equitable access to knowledge in the AI era?

G. Literature Review

The intersection between artificial intelligence and copyright law has become an increasingly significant topic in contemporary legal scholarship. With the emergence of generative AI systems capable of producing human-like creative works, scholars have begun examining the implications of machine learning technologies on traditional copyright frameworks. Existing literature broadly focuses on three major themes: the legality of AI training datasets, the political economy of academic publishing, and the adequacy of international intellectual property regimes in addressing emerging technological challenges.

H. Research Methodology

1. Doctrinal Legal Research Approach

This study adopts a doctrinal legal research methodology to analyze the evolving relationship between artificial intelligence technologies and copyright law. Doctrinal research primarily involves the systematic examination of legal principles, statutes, judicial decisions, and international treaties. By analyzing existing legal frameworks, the study seeks to evaluate whether current copyright regimes adequately address the challenges posed by artificial intelligence and large-scale data extraction.

2. Use of Secondary Sources

The research relies extensively on secondary sources including academic journals, legal commentaries, policy reports, and conference publications. These materials provide critical insights into ongoing scholarly debates surrounding AI governance, intellectual property protection, and digital knowledge ecosystems.

Additionally, reports published by international organizations such as the World Intellectual Property Organization and global policy discussions held during international AI forums such as the AI for Good Global Summit have been examined to understand contemporary regulatory trends.

3. Treaty and Policy Analysis

A significant portion of this research involves analyzing international intellectual property treaties and policy frameworks. Particular attention is given to the provisions of the TRIPS Agreement, which establishes minimum standards for copyright protection among member states.

The study evaluates whether existing treaty provisions especially those related to copyright limitations and exceptions can accommodate emerging technologies such as artificial intelligence and machine learning systems.

4. Comparative Legal Perspective

Where relevant, the research adopts a comparative legal perspective to examine how different jurisdictions approach the regulation of AI training datasets. Several countries have begun introducing text-and-data mining exceptions or AI governance frameworks to address the legal uncertainties surrounding machine learning technologies. By comparing these regulatory approaches, the paper aims to identify potential policy solutions that could balance technological innovation with the protection of intellectual property rights.

IV. ARTIFICIAL INTELLIGENCE AND COPYRIGHT SCHOLARSHIP

Legal scholars have extensively debated whether the use of copyrighted materials in AI training constitutes copyright infringement or whether it should fall within existing

copyright exceptions such as fair use or text-and-data mining provisions. Some researchers argue that machine learning processes involve the reproduction of copyrighted material during dataset preparation and model training, thereby triggering copyright protections. Others contend that the use of such materials is transformative because the purpose of machine learning is to extract patterns rather than reproduce expressive content.[8]

This debate has intensified with the development of generative AI models that can produce outputs resembling existing creative works. Critics argue that these systems may indirectly replicate the expressive elements of copyrighted works contained in their training datasets, thereby raising concerns about unauthorized derivative creation. At the same time, proponents of AI innovation emphasize the societal benefits of machine learning technologies and argue that overly restrictive copyright rules could hinder technological progress.[9]

A. Fair Use and Text-and-Data Mining

The concept of fair use plays a central role in debates surrounding AI training datasets. Under certain legal frameworks, particularly in the United States, courts have recognized that the use of copyrighted works for purposes such as data analysis or search indexing may qualify as transformative use. Several scholars draw parallels between AI training processes and earlier digital technologies such as search engines and data indexing systems, which were ultimately recognized as legitimate uses under copyright law.[10]

However, the legal applicability of fair use in the context of AI remains uncertain. Machine learning models often require the ingestion of enormous datasets consisting of millions of documents, images, and audiovisual materials. This large-scale extraction of copyrighted content may exceed the scope of traditional copyright exceptions,

[8] Gervais, D. (2022). Artificial intelligence and copyright: Ownership and authorship issues. *Vanderbilt Journal of Entertainment & Technology Law*, 24(3).
[9] Abbott, R. (2020). *The Reasonable Robot: Artificial Intelligence and the Law*. Cambridge University Press.
[10] Leval, P. (1990). Toward a fair use standard. *Harvard Law Review*, 103.

particularly in jurisdictions where fair use doctrines are narrower or where text-and-data mining exceptions are limited to non-commercial research purposes.[11]

B. Academic Publishing and Knowledge Monopolies

Another major theme within the literature concerns the concentration of power within the academic publishing industry. Scholars have long criticized the subscription-based publishing model in which major academic publishers control access to scholarly knowledge through expensive journal subscriptions. This model has been widely criticized for restricting access to publicly funded research while generating significant profits for commercial publishers.[12]

Large publishing houses such as Elsevier, Springer Nature, and Wiley dominate the global academic publishing market and maintain extensive digital databases that store millions of research articles. Access to these databases is typically restricted to institutions capable of paying high subscription fees, thereby limiting the availability of scholarly knowledge for independent researchers and institutions in developing countries.

Ironically, while human researchers often face these barriers, artificial intelligence systems may still access large portions of online academic literature through web scraping and digital data collection techniques. This contradiction has been described by some scholars as the "AI knowledge paradox," where machine learning systems benefit from access to knowledge that remains inaccessible to many human researchers.

C. Global Intellectual Property Governance

The role of international intellectual property law has also been widely discussed in academic literature. Global copyright standards are primarily governed by international treaties that establish minimum protection requirements for member states. The most influential of these agreements is the TRIPS Agreement, which operates within the institutional framework of the World Trade Organization.

---

[11] Margoni, T., & Kretschmer, M. (2018). The text and data mining exception in the EU copyright directive. *European Intellectual Property Review*.
[12] Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers. *PLOS ONE*.

However, scholars frequently argue that these international agreements were designed before the emergence of modern digital technologies and therefore struggle to address contemporary issues related to artificial intelligence, big data, and digital platforms. As a result, there is growing academic consensus that international copyright frameworks must evolve to address the technological realities of the AI era.[13]

D. Research Gap

Despite the growing body of scholarship on artificial intelligence and copyright law, relatively few studies have examined the combined relationship between AI training datasets, academic publishing paywalls, and international copyright governance. Most research focuses either on copyright law or on the economics of academic publishing, without analyzing how these issues interact within the broader global knowledge ecosystem.

Furthermore, the concept of "copyright capture" remains underexplored in legal literature, particularly in relation to the economic advantages enjoyed by large technology companies. This paper seeks to address this gap by examining how AI training practices, corporate technological power, and international intellectual property regimes interact to shape the future of knowledge access and copyright governance.

V. COPYRIGHT LAW IN THE AGE OF ARTIFICIAL INTELLIGENCE

A. Traditional Copyright Principles

Copyright law has historically been designed to protect original creative expressions produced by human authors. Core copyright principles include originality, authorship, fixation, and exclusive rights granted to the creator of a work. These exclusive rights typically include the right to reproduce, distribute, adapt, publicly perform, and display the work.[14]

---

[13] Yu, P. K. (2020). Intellectual property and global governance in the digital age. *Michigan State Law Review*.
[14] Goldstein, P., & Hugenholtz, B. (2019). *International Copyright: Principles, Law, and Practice*. Oxford University Press.

Under most copyright regimes, protection is automatically granted once a work is expressed in a tangible medium. The purpose of this legal framework is to encourage creativity and innovation by granting creators a limited monopoly over the use of their works while ultimately ensuring that knowledge and creative expression benefit society as a whole.[15]

However, the emergence of artificial intelligence technologies challenges several foundational assumptions of copyright law. Traditional copyright doctrines were developed in a context where creative works were produced exclusively by human authors. AI systems, on the other hand, operate through automated processes that analyze vast quantities of existing data in order to generate new outputs. This development raises important questions regarding the application of traditional copyright principles to machine learning processes.

B. Reproduction Rights in Digital Environments

One of the most significant copyright issues raised by artificial intelligence concerns the reproduction right. AI training processes typically involve copying large quantities of digital content into datasets that are then used to train machine learning models. During this process, copyrighted materials such as books, research articles, images, and audio recordings may be temporarily reproduced within the computational environment.[16]

From a legal perspective, such reproductions may fall within the exclusive rights of copyright holders. In traditional contexts, reproducing copyrighted works without authorization would generally constitute copyright infringement unless it falls within an applicable exception. The scale of copying involved in AI training processes therefore raises concerns about the legality of such practices.

The complexity of this issue becomes even more pronounced when AI developers collect data from the internet using automated scraping technologies. These practices allow

[15] Landes, W., & Posner, R. (2003). *The Economic Structure of Intellectual Property Law*. Harvard University Press.
[16] Guadamuz, A. (2017). Artificial intelligence and copyright. *WIPO Magazine*.

developers to gather vast datasets from publicly accessible websites, many of which contain copyrighted materials.

Large technology corporations including Google and Microsoft possess the computational infrastructure necessary to process enormous datasets consisting of billions of digital documents. The availability of such resources creates significant disparities between large technology firms and smaller developers, further intensifying debates regarding data ownership and copyright governance.

C. Copyright Implications of AI Training Datasets

Modern AI models rely heavily on extensive training datasets to develop linguistic and analytical capabilities. These datasets may include digital books, academic journals, websites, and other forms of copyrighted material. Because many of these works remain protected under copyright law, their inclusion within training datasets raises questions about whether AI developers must obtain licenses from copyright holders.

Supporters of AI innovation argue that training algorithms on large datasets does not involve the expressive use of copyrighted works but rather involves statistical analysis of patterns and relationships within the data. According to this argument, the purpose of machine learning is not to reproduce or distribute the original works but to enable computers to learn from them.[17]

However, critics argue that AI systems may generate outputs that replicate or closely resemble portions of the copyrighted works used in training. If such outputs contain identifiable elements of the original works, copyright infringement claims may arise.

The debate has intensified with the rise of generative AI systems developed by companies such as OpenAI. These systems can produce text, images, and other creative outputs that appear similar to human-generated works, raising concerns about authorship, originality, and the potential displacement of human creators.

[17] Lemley, M. A., & Casey, B. (2021). Fair learning. *Texas Law Review*, 99.

D. Fair Use and Text-and-Data Mining Exceptions

One of the primary legal arguments advanced by AI developers is that the use of copyrighted materials for machine learning constitutes fair use or an equivalent copyright exception. In certain jurisdictions, courts have recognized that copying copyrighted works for purposes such as search indexing, data analysis, or research may qualify as transformative use.

The concept of text-and-data mining has become particularly relevant in this context. Text-and-data mining refers to the automated analysis of large datasets in order to extract patterns, relationships, or insights. Some legal scholars argue that AI training should be treated similarly to other forms of data analysis that have historically been permitted under copyright law.[18]

Nevertheless, the legal recognition of such exceptions varies across jurisdictions. While some countries have introduced specific text-and-data mining exceptions, others maintain stricter copyright regimes that limit the scope of permissible data use.

This lack of uniformity creates significant uncertainty for AI developers operating in a global digital environment.

E. AI-Generated Outputs and the Question of Authorship

Another major challenge posed by artificial intelligence relates to the authorship of AI-generated works. Traditional copyright law recognizes human authors as the creators of protected works. Because AI systems operate autonomously once trained, determining the author of AI-generated content becomes legally complex.

Some scholars argue that copyright protection should be granted to the human developers or users who design and operate AI systems. Others suggest that AI-generated works should remain in the public domain because they lack human authorship.[19]

---

[18] Margoni, T. (2018). The EU text and data mining exception. *European Intellectual Property Review*.
[19] Abbott, R. (2020). *The Reasonable Robot: Artificial Intelligence and the Law*. Cambridge University Press.

Courts and policymakers have increasingly begun to address this issue. Several copyright offices around the world have refused to grant copyright protection to works generated entirely by artificial intelligence without significant human involvement.

F. Emerging Copyright Litigation in the AI Industry

The rapid growth of generative AI technologies has also led to a wave of copyright litigation involving AI developers. Authors, artists, and media organizations have filed lawsuits alleging that AI companies used their copyrighted works to train machine learning models without obtaining permission.

Several legal actions have targeted companies involved in generative AI development, including those associated with platforms developed by OpenAI and other technology firms. These cases raise fundamental questions regarding the legality of AI training datasets and the scope of copyright protections in the digital age.

Although many of these disputes remain unresolved, they are likely to play a critical role in shaping the future legal landscape of artificial intelligence and copyright governance.

VI. BIG TECH AND COPYRIGHT CAPTURE

A. Concept of Copyright Capture

The rapid expansion of artificial intelligence technologies has led to the emergence of a new phenomenon often described in academic literature as "copyright capture." The concept refers to situations where large technology corporations extract economic value from copyrighted works by incorporating them into machine learning datasets without establishing clear compensation mechanisms for the original creators.[20]

Unlike traditional copyright infringement, copyright capture does not necessarily involve direct reproduction or distribution of copyrighted works. Instead, it involves the large-scale ingestion of digital content into AI training systems, allowing algorithms to learn patterns, structures, and stylistic elements from existing works. These learned patterns

[20] Benkler, Y. (2017). Peer production and the commons in the digital economy. *Yale Law Journal*.

then enable AI systems to generate new content that may resemble the works used in training.

This process raises important ethical and legal concerns because the economic benefits generated by AI systems often accrue primarily to technology companies rather than to the authors whose works contributed to the training process.

B. Data Extraction and Web Scraping Practices

Modern artificial intelligence models rely on extremely large datasets containing billions of words, images, and multimedia files. These datasets are frequently assembled through automated data extraction techniques such as web scraping, which allows developers to collect publicly accessible digital content from websites, online forums, digital libraries, and news platforms. The legality of such collection remains contested where scraped materials include copyrighted works used for machine learning training, particularly in relation to reproduction rights, fair use, and licensing markets.[21]

Web scraping technologies allow AI developers to gather data at an unprecedented scale. While many of these sources are publicly available online, they often contain copyrighted materials that remain protected under national copyright laws. Consequently, the legality of using such materials for AI training purposes has become a central issue in ongoing legal and policy debates.

Technology companies argue that the use of publicly available internet data for machine learning purposes constitutes legitimate data analysis. Critics, however, contend that large-scale scraping of copyrighted materials represents a form of digital exploitation that undermines the rights of creators.

C. Economic Advantages of Large Technology Corporations

The development of advanced AI systems requires enormous computational resources, specialized expertise, and access to large datasets. These requirements create significant barriers to entry for smaller companies and research institutions. As a result, the global

[21] Bridy, A. (2022). Artificial intelligence and copyright data scraping. *Stanford Technology Law Review*.

AI industry has become increasingly concentrated among a small number of major technology corporations.

Companies such as Google, Microsoft, and OpenAI possess the financial resources, cloud computing infrastructure, and engineering talent necessary to develop large-scale machine learning models. Their access to vast digital datasets provides them with a substantial competitive advantage in the development of generative AI technologies.

This concentration of technological power has led some scholars to describe the emerging AI economy as a form of "data capitalism," where control over large datasets becomes a key source of economic and political influence.[22]

D. AI Training and the Knowledge Economy

Artificial intelligence systems rely heavily on data extracted from a wide range of digital sources, including books, academic publications, news articles, social media posts, and multimedia content. Many of these materials represent the intellectual labor of writers, journalists, researchers, artists, and educators.

When such materials are incorporated into AI training datasets without compensation or consent, concerns arise regarding the fairness of the resulting economic model. Technology companies may generate substantial revenues from AI products built upon knowledge created by others.

This dynamic becomes particularly problematic when considering the broader knowledge economy. Academic research and creative works often receive public funding or institutional support. If private corporations are able to commercialize AI systems trained on these materials without sharing benefits with the original creators, the sustainability of knowledge production may be undermined.

[22] Srnicek, N. (2017). *Platform Capitalism*. Polity Press

E. Emerging Copyright Litigation Against AI Developers

The growing controversy surrounding AI training datasets has resulted in several significant copyright disputes involving authors, publishers, news organizations, and technology companies. These cases are important because they directly test whether the use of copyrighted works for AI training can be justified as fair use, lawful data analysis, or whether it amounts to unauthorized reproduction requiring permission or licensing.

In *The New York Times Co. v. Microsoft Corp. and OpenAI*, filed in the Southern District of New York on 27 December 2023, the plaintiff alleged that OpenAI and Microsoft used large quantities of copyrighted journalistic content to train and operate generative AI systems. The dispute is significant because it concerns both the training of large language models and the alleged generation of outputs that reproduce or closely summarize protected news content. It therefore directly raises questions of market substitution, licensing harm, and the limits of fair use in relation to journalistic works.

In *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc.*, the District of Delaware rejected Ross Intelligence's fair use defence in February 2025 in relation to the use of Westlaw headnotes and legal research content for developing an AI-powered legal research tool. The decision is doctrinally important because it treated the defendant's use as insufficiently transformative where the resulting product competed with the plaintiff's legal research market. It also demonstrates that courts may give substantial weight to the effect of AI training uses on existing or potential licensing markets.

In *Bartz v. Anthropic PBC*, authors challenged Anthropic's use of copyrighted books in connection with large language model development. The litigation became especially significant after a major class settlement of approximately USD 1.5 billion concerning pirated book copies. The case illustrates an emerging judicial distinction between AI training on lawfully acquired works, which may be argued as transformative in some circumstances, and the unauthorized acquisition or retention of pirated datasets, which presents a more serious infringement concern.

Together, these disputes show that AI-copyright litigation is no longer speculative. It now directly shapes the legal boundaries of training datasets, fair use, licensing markets, and the responsibility of AI developers to obtain authorization where copyrighted works are used in ways that affect authors, publishers, and knowledge producers.[23]

F. Copyright Capture and Global Inequality

The phenomenon of copyright capture also raises broader concerns about global inequality within the digital knowledge ecosystem. Many developing countries contribute large amounts of cultural and intellectual content to the global internet, yet they often lack the technological infrastructure required to develop advanced AI systems.

As a result, technology companies based in technologically advanced economies may benefit disproportionately from the use of globally sourced data. This dynamic may reinforce existing disparities between developed and developing countries in terms of technological capabilities, economic benefits, and access to knowledge resources.

The issue has increasingly become part of global policy discussions within international organizations such as the World Intellectual Property Organization and international AI governance forums, including the India AI Impact Summit 2026. This summit formed part of the continuing international AI governance sequence following the Bletchley Park AI Safety Summit 2023, the AI Seoul Summit 2024, and the AI Action Summit 2025 in Paris.

These forums have begun examining how international intellectual property frameworks should evolve in response to artificial intelligence technologies and the growing concentration of data power among large technology corporations.

[23] Ginsburg, J. (2023). Generative AI and copyright law. *Columbia Journal of Law & the Arts*.

VII. ACADEMIC PAYWALLS AND THE POLITICAL ECONOMY OF KNOWLEDGE

A. Structure of the Academic Publishing Industry

The modern academic publishing industry plays a central role in the global dissemination of scholarly knowledge. Academic journals serve as primary platforms through which researchers communicate findings, contribute to scholarly debates, and advance scientific and intellectual progress. However, the structure of the academic publishing market has been widely criticized for its heavy concentration among a small number of large commercial publishers.[24]

These publishers manage thousands of academic journals across multiple disciplines and maintain extensive digital databases containing millions of research articles. Access to these databases is typically provided through institutional subscriptions purchased by universities, research institutes, and libraries. Because subscription costs are often extremely high, many institutions—particularly those located in developing countries—struggle to obtain access to essential academic resources.

This economic structure has created significant inequalities in the global distribution of knowledge. While researchers at well-funded institutions enjoy extensive access to scholarly materials, independent scholars and institutions with limited financial resources face substantial barriers to accessing the same information.

B. Paywall-Based Journal Access Models

The dominant business model within academic publishing relies on subscription-based access systems commonly referred to as paywalls. Under this model, readers must either purchase individual articles or obtain access through institutional subscriptions in order to read scholarly publications.[25]

---

[24] Fyfe, A. et al. (2017). *Untangling Academic Publishing: A History of the Relationship Between Commercial Interests, Academic Prestige and the Circulation of Research*. University of St Andrews.
[25] Suber, P. (2012). *Open Access*. MIT Press.

These paywalls restrict the availability of academic knowledge even though much of the underlying research is publicly funded through government grants, university resources, and taxpayer-supported research programs. Critics argue that the paywall model creates a paradox in which publicly funded knowledge becomes privately controlled by commercial publishers.

At the same time, digital technologies have made it technically possible to distribute academic research more widely than ever before. The internet has dramatically reduced the costs associated with publishing and distributing scholarly works. Despite this technological transformation, the economic structure of academic publishing has remained largely unchanged.

C. Dominance of Major Academic Publishers

The global academic publishing industry is dominated by a small group of multinational publishing corporations that control a significant portion of the scholarly journal market. Major publishers include:

1. Elsevier
2. Springer Nature
3. Wiley

These companies operate extensive digital platforms that host academic journals, conference proceedings, and research databases. Their subscription packages are often sold to universities as bundled collections of journals, which can cost institutions millions of dollars annually.

Scholars have frequently described this structure as an oligopoly, in which a small number of firms exercise significant control over the global flow of academic information.[26] This concentration of market power has enabled major publishers to

[26] Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLOS ONE*.

maintain high profit margins while continuing to rely heavily on unpaid academic labor for peer review and editorial activities.

D. Impact on Researchers and Academic Institutions

The paywall-based publishing system has profound consequences for researchers and academic institutions around the world. Universities often allocate substantial portions of their library budgets to journal subscriptions, leaving fewer resources available for other research activities.[27]

For researchers working outside major academic institutions, access barriers can be even more severe. Independent scholars, practitioners, and policymakers may find it difficult to obtain access to scholarly research unless they are affiliated with institutions capable of paying subscription fees.

This situation has raised serious concerns about the inclusivity and accessibility of global knowledge systems. In many developing countries, universities operate with limited research budgets and cannot afford access to major academic databases. As a result, researchers may face significant obstacles in accessing the latest scientific and legal scholarship.

E. The Open Access Movement

In response to growing criticism of paywall-based publishing models, the open access movement has emerged as a major reform initiative within academic publishing. Open access advocates argue that scholarly research should be freely accessible to readers without subscription barriers.[28]

Several alternative publishing models have been proposed to support open access distribution. These include institutional repositories, publicly funded open access journals, and author-funded publication models. Many governments and funding

[27] Eve, M. (2014). *Open Access and the Humanities*. Cambridge University Press.
[28] Budapest Open Access Initiative. (2002). Open access declaration.

agencies have also begun requiring that publicly funded research be made openly available.

Despite these efforts, the transition toward open access publishing has been gradual and uneven. Many high-impact journals continue to operate under traditional subscription models, and authors may face high article processing charges in some open access journals.

F. The AI Knowledge Paradox

The rise of artificial intelligence technologies has introduced a new dimension to debates about academic publishing and knowledge access. AI developers frequently train machine learning models using vast collections of digital text, including academic articles and books that may be subject to copyright protection or subscription access restrictions.[29]

This situation has created what some scholars describe as the AI knowledge paradox. While human researchers may face barriers in accessing scholarly literature due to paywalls, AI systems may still be able to analyze large volumes of online academic content through automated data collection techniques.

This paradox highlights the broader tension between the commercialization of academic knowledge and the technological capacity to distribute information widely in the digital era. If artificial intelligence systems can benefit from access to global knowledge resources while many human researchers remain excluded, the legitimacy of existing publishing models may increasingly come under scrutiny.

Furthermore, the interaction between AI training datasets and academic publishing systems raises important questions regarding copyright ownership, licensing frameworks, and equitable access to knowledge in the digital age.

[29] Tennant, J. (2020). The evolving open access ecosystem. *Learned Publishing*.

VIII. THE TRIPS DILEMMA IN THE AI ERA

A. Overview of the TRIPS Framework

International intellectual property regulation is largely shaped by multilateral treaties that establish minimum protection standards for member states. Among these agreements, the most influential is the TRIPS Agreement, which forms an integral part of the legal framework governing global trade under the World Trade Organization.

Adopted in 1994 as part of the Uruguay Round of trade negotiations, the TRIPS Agreement introduced comprehensive international standards for the protection and enforcement of intellectual property rights. These standards apply across several areas of intellectual property law, including copyright, trademarks, patents, and trade secrets.[30]

The primary objective of the TRIPS Agreement is to balance the protection of intellectual property with the promotion of technological innovation and the transfer of knowledge. By establishing uniform minimum standards, the treaty seeks to reduce disparities between national intellectual property regimes and facilitate international trade in knowledge-based goods and services.

However, the agreement was negotiated during a period when digital technologies and artificial intelligence had not yet become central components of the global knowledge economy. Consequently, many of its provisions do not directly address issues related to machine learning, automated data extraction, or AI-generated content.

B. Copyright Standards under TRIPS

The TRIPS Agreement incorporates several provisions relating to copyright protection, primarily by integrating key principles from earlier international copyright treaties such as the Berne Convention. Under TRIPS, member states are required to provide copyright

[30] Correa, C. (2007). *Trade Related Aspects of Intellectual Property Rights: A Commentary on the TRIPS Agreement*. Oxford University Press.

protection for literary and artistic works, including books, articles, computer programs, and other creative expressions.[31]

The agreement also requires member states to recognize the exclusive rights of copyright holders, including reproduction rights, distribution rights, and rights related to public communication of works. These rights are designed to ensure that creators maintain control over the commercial exploitation of their works.

In the context of artificial intelligence, these provisions become particularly relevant because AI training processes frequently involve the reproduction and analysis of copyrighted materials. If such reproductions occur without authorization from copyright holders, questions may arise regarding compliance with international copyright standards established under TRIPS.

C. Article 13 and the Three-Step Test

One of the most important provisions of the TRIPS Agreement in relation to copyright exceptions is Article 13, which establishes what is commonly referred to as the three-step test. This provision allows member states to introduce limitations or exceptions to copyright protection under certain conditions.

According to the three-step test, copyright exceptions must satisfy three criteria:

1. They must be confined to certain special cases.
2. They must not conflict with the normal exploitation of the work.
3. They must not unreasonably prejudice the legitimate interests of the copyright holder.[32]

This test serves as a guiding framework for evaluating whether specific copyright exceptions such as those related to research, education, or data analysis—are consistent with international copyright law.

[31] Gervais, D. (2012). *The TRIPS Agreement: Drafting History and Analysis*. Sweet & Maxwell.
[32] Senftleben, M. (2004). Copyright, limitations and the three-step test. *Kluwer Law International*

In the context of artificial intelligence, policymakers must consider whether activities such as text-and-data mining or machine learning training can be accommodated within the scope of these exceptions. The challenge lies in determining whether the large-scale data extraction required for AI development can be considered a "special case" under the three-step test.

D. Limitations of TRIPS in Addressing Artificial Intelligence

Although the TRIPS Agreement remains one of the most important international instruments governing intellectual property rights, it was not designed to address the complexities introduced by modern digital technologies. Artificial intelligence systems rely on automated data processing techniques that were not contemplated during the negotiation of the treaty.

As a result, the application of TRIPS provisions to AI training datasets often involves significant interpretative challenges. For example, the agreement does not explicitly address whether the automated copying of digital content for machine learning purposes constitutes copyright infringement or whether it may qualify as a permissible research activity.[33]

Additionally, the global nature of AI development complicates the application of national copyright laws. Machine learning models may be trained using data collected from multiple jurisdictions, each with its own copyright regulations and exceptions. This creates legal uncertainty for developers seeking to operate within international legal frameworks.

E. Cross-Border Issues in AI Training Data

Artificial intelligence development is inherently transnational in nature. Training datasets often include materials sourced from websites, digital libraries, and media platforms located across different countries. As a result, AI developers may inadvertently incorporate copyrighted works originating from multiple jurisdictions.

[33] Yu, P. K. (2017). Intellectual property and the information economy. *Michigan State Law Review*.

This cross-border dimension raises complex legal questions regarding jurisdiction, enforcement, and applicable copyright standards. If a dataset contains copyrighted materials from multiple countries, determining which legal framework applies can become extremely challenging.

Furthermore, the uneven distribution of technological resources across the world may lead to asymmetries in the benefits derived from AI development. Technology companies located in technologically advanced economies may utilize globally sourced data to develop profitable AI systems, while creators in other parts of the world may receive little or no compensation.

F. Implications for Developing Countries

The challenges posed by artificial intelligence technologies are particularly significant for developing countries. Many developing nations contribute valuable cultural and intellectual content to the global digital ecosystem, yet they often lack the technological infrastructure necessary to develop advanced AI systems.

This situation raises concerns regarding the equitable distribution of benefits derived from AI innovation. If large technology companies are able to utilize globally sourced datasets without appropriate regulatory oversight, the resulting economic benefits may become concentrated within a small number of technologically advanced economies.

policy forums, including discussions organized by the World Intellectual Property Organization and global technology governance platforms such as the India AI Impact Summit 2026. The 2026 New Delhi summit continued the broader international dialogue shaped by the Bletchley Park AI Safety Summit 2023, the AI Seoul Summit 2024, and the AI Action Summit 2025 in Paris.

Such forums aim to explore potential regulatory frameworks that can ensure both technological innovation and equitable access to knowledge resources in the emerging AI-driven global economy.

IX. GLOBAL AI GOVERNANCE AND INTERNATIONAL POLICY DEBATES

A. Emergence of Global Artificial Intelligence Governance

The rapid advancement of artificial intelligence technologies has generated increasing concern among policymakers, legal scholars, and international organizations regarding the need for global regulatory frameworks. Artificial intelligence is not confined within national boundaries; rather, it operates within a highly interconnected digital ecosystem in which data flows across jurisdictions and AI systems are deployed globally. As a result, national regulatory approaches alone are often insufficient to address the legal, ethical, and economic challenges posed by AI technologies.[34]

Governments and international institutions have therefore begun exploring mechanisms for international cooperation in AI governance. These initiatives seek to balance the promotion of technological innovation with the protection of fundamental rights, intellectual property, and equitable access to knowledge resources.

The debate surrounding AI governance has become particularly significant in the context of generative AI systems capable of producing text, images, and multimedia outputs based on large-scale training datasets. Such systems raise questions regarding copyright ownership, accountability, transparency, and the ethical use of data.

B. Policy Dialogues within International Organizations

International organizations have played an important role in facilitating discussions on AI governance and intellectual property regulation. The World Intellectual Property Organization has emerged as a key platform for examining the relationship between artificial intelligence and intellectual property rights. Through expert consultations and policy dialogues, the organization has been exploring issues related to AI-generated

[34] Floridi, L. et al. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*.

works, data ownership, and the implications of machine learning for existing copyright frameworks.[35]

These discussions have highlighted the need to reconsider traditional intellectual property doctrines in light of technological developments. In particular, policymakers have emphasized the importance of ensuring that copyright regimes remain flexible enough to accommodate emerging technologies while continuing to protect the legitimate interests of creators.

Similarly, international organizations involved in digital governance and technology policy have recognized the need for coordinated global responses to AI-related challenges. These discussions often focus on ethical principles, transparency standards, and regulatory mechanisms designed to promote responsible AI development.

C. Ethical AI Frameworks and UNESCO Initiatives

Ethical considerations have become a central component of global AI governance debates. Concerns regarding algorithmic bias, transparency, accountability, and human oversight have prompted policymakers to develop ethical frameworks intended to guide the development and deployment of AI technologies.

One of the most prominent initiatives in this area has been led by UNESCO, which has developed global recommendations on the ethical use of artificial intelligence. These recommendations emphasize principles such as fairness, inclusivity, transparency, and respect for human rights in the development of AI systems.[36]

Within the context of copyright and knowledge access, these ethical frameworks highlight the importance of ensuring that AI technologies do not exacerbate existing inequalities within global knowledge systems. If AI development relies heavily on data extracted from creative and academic works, policymakers must consider how the resulting economic benefits are distributed among stakeholders.

---

[35] World Intellectual Property Organization. (2020). *WIPO Conversation on Intellectual Property and Artificial Intelligence*.
[36] UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*.

D. The EU AI Act and Copyright Transparency Obligations

A significant recent development in global AI governance is the European Union's Artificial Intelligence Act, formally adopted as Regulation (EU) 2024/1689. Although the Act is primarily structured as a risk-based framework for regulating artificial intelligence, it is directly relevant to copyright governance because it imposes specific obligations on providers of general-purpose AI models. Article 53 requires such providers to maintain a policy for compliance with Union copyright law, including respect for rights reservations under the text-and-data mining framework of Article 4(3) of the Digital Single Market Directive. It also requires providers to publish a sufficiently detailed summary of the content used for training their models, according to a template prepared by the AI Office.

This development is particularly important for the present study because it reflects a concrete legislative response to the problem of opaque AI training datasets. Unlike general ethical principles or non-binding policy declarations, Article 53 creates a binding transparency mechanism that may assist copyright holders, researchers, and regulators in understanding the categories of materials used in AI training. At the same time, the EU model also reveals practical enforcement challenges. The obligation requires a sufficiently detailed summary rather than full disclosure of training datasets, and its effectiveness will depend on the quality of the template, regulatory supervision, and the ability of copyright holders to identify whether their works were used. Accordingly, the EU AI Act provides a useful comparative case study for evaluating both the promise and limitations of dataset-transparency regulation in the AI era.

E. International Conferences and AI Policy Summits

In recent years, global conferences and policy summits have become important venues for discussing the regulatory challenges posed by artificial intelligence. These events bring together government officials, technology companies, academics, and civil society organizations to explore potential governance frameworks.

One significant forum for such discussions is the AI for Good Global Summit, which focuses on the use of artificial intelligence technologies to address global challenges such as climate change, healthcare access, and sustainable development. The summit provides a platform for dialogue on how AI can be used responsibly while minimizing potential risks.

Similarly, international policy debates were advanced during the India AI Impact Summit 2026, held from 16–20 February 2026 at Bharat Mandapam, New Delhi. The summit should be situated within the continuing chronology of global AI governance forums, following the Bletchley Park AI Safety Summit 2023, the AI Seoul Summit 2024, and the AI Action Summit 2025 in Paris. These discussions included topics such as AI transparency, data governance, intellectual property protection, responsible innovation, and the economic implications of generative AI technologies.

Such global policy dialogues demonstrate growing recognition that artificial intelligence governance requires international cooperation. Because AI technologies rely on globally sourced datasets and operate across multiple jurisdictions, effective regulatory frameworks must address cross-border legal challenges.

F. Implications for the Global South

The rise of artificial intelligence technologies also has important implications for developing countries, often collectively referred to as the Global South. Many developing countries possess rich cultural, linguistic, and intellectual resources that contribute to the global digital ecosystem. However, they frequently lack the technological infrastructure necessary to develop advanced AI systems at the same scale as major technology corporations.

As a result, there is growing concern that AI development may reinforce existing inequalities in the global knowledge economy. If large technology companies are able to extract data from globally available sources while retaining exclusive control over AI technologies, developing countries may remain dependent on foreign technological infrastructure.

These concerns have been increasingly highlighted in international policy discussions. Policymakers have emphasized the importance of ensuring that AI governance frameworks promote inclusive technological development and equitable access to knowledge resources.

X. POLICY RECOMMENDATIONS

The preceding discussion highlights the complex intersection between artificial intelligence technologies, copyright protection, academic publishing systems, and international intellectual property law. As AI systems continue to evolve, policymakers must develop regulatory frameworks that balance the interests of technological innovation, knowledge accessibility, and the protection of creators' rights. This section proposes several policy recommendations aimed at addressing the challenges associated with copyright capture in the AI age.

A. Regulation of AI Training Datasets

One of the most pressing issues in the governance of artificial intelligence concerns the use of copyrighted materials in training datasets. Many AI developers rely on large-scale web scraping and automated data collection techniques to obtain digital content from online sources. While such practices facilitate technological innovation, they also raise concerns about unauthorized reproduction, market substitution, licensing obligations, and the need for transparent governance of AI training datasets.[37]

Policymakers should therefore consider introducing clearer regulatory frameworks governing the use of copyrighted materials in AI training processes. One possible approach is the establishment of transparent dataset disclosure requirements, under which AI developers would be required to disclose sufficiently detailed summaries of the categories and sources of content used to train their models. This recommendation is no longer merely theoretical, as Article 53 of the EU AI Act already requires providers of general-purpose AI models to maintain copyright-compliance policies and publish

[37] Bridy, A. (2022). Copyright in the age of artificial intelligence. *Stanford Technology Law Review*.

sufficiently detailed summaries of training content in accordance with an AI Office template.

However, the EU framework should be treated as a starting point rather than a complete solution. Transparency summaries may assist copyright holders in assessing whether protected works were used, but they may not reveal the full dataset or resolve questions of licensing, compensation, or cross-border enforcement. Regulators should therefore combine transparency obligations with fair licensing frameworks that allow creators to grant permission for AI training under clear, equitable, and enforceable condition.

B. Licensing Frameworks for AI Training Data

Another potential policy solution involves the development of collective licensing mechanisms for AI training datasets. Similar to existing systems used in music copyright management, collective licensing organizations could facilitate the licensing of large bodies of copyrighted works for use in machine learning training processes.

Such frameworks would allow AI developers to obtain legal authorization for the use of copyrighted materials while ensuring that creators receive appropriate compensation for the use of their works.[38]

Collective licensing models may be particularly useful in cases where AI training datasets include millions of individual works, making direct negotiations with individual copyright holders impractical.

C. Strengthening Open Access Publishing Models

Reforming academic publishing systems is another important step in addressing the challenges associated with copyright capture. The widespread use of paywalls within academic publishing restricts access to scholarly knowledge and creates barriers for researchers, particularly those located in developing countries.

[38] Kretschmer, M., & Margoni, T. (2021). Licensing models for AI training data. *Journal of Intellectual Property Law & Practice*.

Encouraging the expansion of open access publishing models can help reduce these barriers and promote a more inclusive global knowledge ecosystem. Governments and research funding agencies may play an important role in this transition by requiring publicly funded research to be made freely accessible.

Open access policies can ensure that academic knowledge remains available to researchers, educators, and policymakers worldwide while reducing dependence on subscription-based publishing models dominated by major corporations such as Elsevier, Springer Nature, and Wiley.

D. Reforming International Intellectual Property Frameworks

At the international level, the growing influence of artificial intelligence technologies highlights the need to reconsider existing intellectual property frameworks. The TRIPS Agreement, administered by the World Trade Organization, was negotiated during a period when digital technologies and machine learning systems were not yet central to the global economy.

Consequently, the treaty provides limited guidance regarding issues such as AI training datasets, automated data analysis, and AI-generated content. Policymakers may therefore need to explore potential reforms that address the unique challenges posed by artificial intelligence.

Possible reforms could include the clarification of copyright exceptions related to text-and-data mining, as well as the development of international guidelines for the ethical and lawful use of copyrighted materials in AI development.

E. Promoting Equitable Global AI Governance

Finally, international AI governance frameworks should emphasize inclusivity and equitable access to technological benefits. Global policy discussions within organizations such as the World Intellectual Property Organization and international forums such as the AI for Good Global Summit and the AI Action Summit 2026 have highlighted the need for cooperative approaches to AI regulation.

These discussions emphasize that the benefits of artificial intelligence should be shared broadly across societies rather than concentrated within a small number of powerful technology companies. By promoting transparent data governance, fair licensing frameworks, and open access knowledge systems, policymakers can help ensure that AI technologies contribute to a more equitable global knowledge economy.

XI. CONCLUSION

Artificial intelligence technologies are rapidly transforming the global knowledge economy. The development of powerful machine learning systems capable of generating text, images, and multimedia content has introduced significant legal and ethical challenges for existing copyright frameworks. As AI developers increasingly rely on vast datasets derived from online content, the issue of copyright capture has become a central concern in contemporary intellectual property debates.

This paper examined the intersection between artificial intelligence development, academic publishing paywalls, and international copyright governance. The analysis demonstrated that the concentration of technological resources among major corporations—including companies such as Google, Microsoft, and OpenAI—has enabled these entities to extract economic value from globally sourced datasets while raising questions about fairness and accountability.

At the same time, the academic publishing industry continues to operate under subscription-based models that restrict access to scholarly knowledge for many researchers around the world. This contradiction highlights the emergence of an AI knowledge paradox, in which artificial intelligence systems may analyze large volumes of academic literature while human researchers face barriers to accessing the same materials.

The study also examined the limitations of existing international intellectual property frameworks, particularly the TRIPS Agreement, which was developed before the emergence of modern artificial intelligence technologies. While the treaty establishes

important standards for copyright protection, it does not fully address the legal complexities associated with machine learning datasets and AI-generated content.

Global policy discussions within international organizations such as the World Intellectual Property Organization and international AI governance forums demonstrate growing recognition of the need for updated regulatory frameworks. As artificial intelligence continues to reshape the global knowledge landscape, policymakers must develop balanced approaches that protect creators' rights while fostering technological innovation and equitable access to knowledge.

Ultimately, addressing the challenges of copyright capture in the AI age will require coordinated efforts among governments, international organizations, technology companies, and academic institutions. By promoting transparent data governance, fair licensing mechanisms, and inclusive knowledge systems, the global community can ensure that artificial intelligence contributes to a more sustainable and equitable knowledge economy.

XII. BIBLIOGRAPHY

1. Abbott, R. (2020). *The Reasonable Robot: Artificial Intelligence and the Law*. Cambridge University Press.
2. Benkler, Y. (2017). Peer production and the commons in the digital economy. *Yale Law Journal*.
3. Lemley, M. A., & Casey, B. (2021). Fair learning. *Texas Law Review*, 99(4), 743–785.
4. U.S. Copyright Office. (2025). *Copyright and Artificial Intelligence, Part 3: Generative AI Training* (Pre-publication version).
5. Budapest Open Access Initiative. (2002). Open access declaration.
6. Correa, C. (2007). *Trade Related Aspects of Intellectual Property Rights: A Commentary on the TRIPS Agreement*. Oxford University Press.
7. Eve, M. (2014). *Open Access and the Humanities*. Cambridge University Press.
8. Floridi, L. et al. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*.

9. Fyfe, A. et al. (2017). *Untangling Academic Publishing: A History of the Relationship Between Commercial Interests, Academic Prestige and the Circulation of Research*. University of St Andrews.
10. Gervais, D. (2012). *The TRIPS Agreement: Drafting History and Analysis*. Sweet & Maxwell.
11. Gervais, D. (2022). Artificial intelligence and copyright: Ownership and authorship issues. *Vanderbilt Journal of Entertainment & Technology Law*, 24(3).
12. Ginsburg, J. (2023). Generative AI and copyright law. *Columbia Journal of Law & the Arts*.
13. Goldstein, P., & Hugenholtz, B. (2019). *International Copyright: Principles, Law, and Practice*. Oxford University Press.
14. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
15. Guadamuz, A. (2017). Artificial intelligence and copyright. *WIPO Magazine*.
16. Kretschmer, M., & Margoni, T. (2021). Licensing models for AI training data. *Journal of Intellectual Property Law & Practice*.
17. Landes, W., & Posner, R. (2003). *The Economic Structure of Intellectual Property Law*. Harvard University Press.
18. Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers. *PLOS ONE*.
19. Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLOS ONE*.
20. Lemley, M. A., & Casey, B. (2021). Fair learning. *Texas Law Review*, 99.
21. Leval, P. (1990). Toward a fair use standard. *Harvard Law Review*, 103.
22. Margoni, T. (2018). The EU text and data mining exception. *European Intellectual Property Review*.
23. Margoni, T., & Kretschmer, M. (2018). The text and data mining exception in the EU copyright directive. *European Intellectual Property Review*.
24. Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux.

25. Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
26. Samuelson, P. (2023). Generative AI meets copyright. *Columbia Journal of Law & the Arts*, 46(2).
27. Senftleben, M. (2004). *Copyright, Limitations and the Three-Step Test*. Kluwer Law International.
28. Srnicek, N. (2017). *Platform Capitalism*. Polity Press.
29. Suber, P. (2012). *Open Access*. MIT Press.
30. Tennant, J. (2020). The evolving open access ecosystem. *Learned Publishing*.
31. UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*.
32. World Intellectual Property Organization. (2020). *WIPO Conversation on Intellectual Property and Artificial Intelligence*.
33. Yu, P. K. (2017). Intellectual property and the information economy. *Michigan State Law Review*.
34. Yu, P. K. (2020). Intellectual property and global governance in the digital age. *Michigan State Law Review*.