ISSN: 2348-2079

International Journal of Intellectual Advancements and Research in Engineering Computations (IJIAREC)

IJIAREC |Vol.13 | Issue 3 | Jul - Sept -2025

www.ijiarec.com

DOI : https://doi.org/10.61096/ijiarec.v13.iss3.2025.143-149

Review

The Intersection of Generative AI and Intellectual Property: Rethinking Fair Dealing for Text and Data Mining in Pharmaceuticals

Dr. Ravi Kumar Kota, K. Mounika*

Pulla Reddy Institute of Pharmacy, Sangareddy-502313, Telangana, India

*Author for Correspondence: K. Mounika
Email: nagasaimounika1234@gmal.com

Published on: 28 Sep 2025

Published by:
Futuristic Publications

Abstract

The rapid advancement of generative artificial intelligence (AI) has revolutionized text and data mining (TDM) practices in the pharmaceutical industry, raising critical questions about intellectual property (IP) protection and the fair dealing doctrine. This article examines the intersection of generative AI and IP, focusing on challenges related to data ownership, licensing, and regulatory compliance. It highlights the need for updated legal frameworks to balance innovation with IP rights, particularly in drug discovery, clinical trials, and personalized medicine. The study provides recommendations for policy reforms, ethical AI use, and international harmonization to support responsible innovation in pharmaceutical regulatory affairs.

Keywords: Generative AI, Intellectual Property, Text and Data Mining, Pharmaceutical Fair Dealing, Regulatory Affairs

INTRODUCTION

Generative artificial intelligence (AI) is transforming the pharmaceutical industry through applications such as automated drug discovery, optimized clinical trial design, and personalized medicine. By leveraging vast datasets, these models can predict molecular structures, analyze clinical outcomes, and tailor treatments to individual patients. However, their heavy reliance on large-scale text and data mining (TDM) often drawing from copyrighted literature, patents, and proprietary databases raises complex intellectual property (IP) challenges. Existing frameworks, including fair dealing provisions that allow limited use of copyrighted material for research, were not designed for AI-generated outputs that blur the distinction between original and derivative works.

This article examines the intersection of generative AI, IP law, and TDM in the pharmaceutical sector, with three core objectives: (1) assess how current legal systems accommodate or conflict with AI-driven

innovation; (2) evaluate regulatory gaps and jurisdictional disparities, particularly in fair dealing provisions and oversight by bodies such as the FDA and EMA; and (3) propose reforms to align IP protection with ethical, legally compliant AI use. Key research questions address the copyrightability and patent eligibility of AI-generated content, the ethical and legal implications of training on copyrighted datasets, and strategies for evolving regulatory frameworks through adaptive fair dealing provisions, AI-specific IP guidelines, and cross-border harmonization[1].

By tackling these issues, this article aims to guide policymakers, industry leaders, and researchers toward a balanced approach that safeguards creators' rights while enabling the transformative potential of generative AI to advance global pharmaceutical innovation.

Generative AI in Pharmaceuticals

Generative AI has emerged as a transformative force in pharmaceutical research and development, demonstrating remarkable capabilities across multiple applications. In drug discovery, AI models like generative adversarial networks (GANs) and variational autoencoders (VAEs) are revolutionizing molecular design by predicting novel compound structures with desired therapeutic properties, significantly accelerating a process that traditionally took years. Clinical trial optimization represents another critical application, where AI algorithms analyze vast datasets to improve patient selection criteria, predict outcomes, and optimize trial protocols, potentially reducing costs and development timelines. Furthermore, AI-powered text and data mining enables comprehensive adverse event detection by processing millions of case reports and scientific literature to identify potential safety signals that might elude manual review. However, these promising applications face significant challenges, particularly concerning data quality and representational bias in training datasets, which can lead to skewed or unreliable outputs. The "black box" nature of many AI systems also raises concerns about transparency and accountability, making it difficult for researchers and regulators to validate AI-generated conclusions or trace decision-making processes. These challenges become particularly acute in highly regulated pharmaceutical applications where patient safety and scientific rigor are paramount[2].

Intellectual Property and Fair Dealing[3]

The intersection of generative AI and intellectual property law presents complex legal questions that current frameworks struggle to address adequately. Copyright law faces particular ambiguity regarding AI-generated content, as traditional concepts of authorship and creativity become blurred when outputs are produced by machines trained on existing copyrighted materials. Central to this debate is the unresolved question of attribution - whether IP rights should vest with the AI developers who created the algorithms, the users who provided input parameters, the data providers whose copyrighted materials were used for training, or whether such outputs should remain in the public domain. The fair dealing doctrine, designed to balance copyright protection with societal needs for access and innovation, shows significant jurisdictional variation in its application to AI and text mining. The United States' flexible fair use doctrine has generally been more accommodating of computational analysis, while the EU's more prescriptive approach under the InfoSoc Directive creates specific but limited exceptions for text and data mining. Other jurisdictions like Canada and Australia take intermediate positions, creating a patchwork of regulations that complicate global research collaborations. These disparities become particularly problematic for pharmaceutical companies operating internationally, where the same AI research activity might be protected in one jurisdiction but constitute infringement in another.

Regulatory Landscape

Pharmaceutical regulators worldwide are grappling with the challenges posed by AI integration into drug development processes. The U.S. Food and Drug Administration (FDA) has taken a proactive stance through its Digital Health Center of Excellence and AI/Machine Learning Action Plan, which outlines a regulatory framework for AI-based medical technologies. Similarly, the European Medicines Agency (EMA) has incorporated AI considerations into its regulatory science strategy, focusing on quality assurance of algorithms and validation of AI-derived evidence. However, these emerging frameworks primarily address clinical applications rather than the IP aspects of AI development. A significant regulatory gap exists in the lack of harmonized standards for protecting AI-generated pharmaceutical innovations, with current IP systems struggling to classify inventions where human involvement is limited to training data provision and parameter setting. The situation is further complicated by the rapid pace of AI advancement, which outstrips the slower evolution of legal and regulatory systems. This misalignment creates uncertainty for pharmaceutical companies investing in AI research and may inadvertently stifle innovation due to fears of IP insecurity or infringement liabilities. The absence of clear international standards also raises challenges for global drug development programs, where consistent IP protection across markets is crucial for securing research investments and bringing new treatments to patients worldwide.

METHODOLOGY

Research Design

This study employs a mixed-methods research approach to comprehensively examine the intersection of generative AI and intellectual property in pharmaceutical applications. The qualitative component involves an in-depth analysis of legal frameworks across key jurisdictions, including copyright laws, patent regulations, and fair dealing exceptions that govern text and data mining practices. Case study methodology forms a crucial part of this qualitative analysis, with detailed examinations of real-world implementations such as Pfizer's AI-driven drug discovery platform, which utilizes machine learning algorithms to identify promising therapeutic compounds. The quantitative dimension complements this through systematic analysis of patent datasets from sources like the USPTO and WIPO, tracking emerging trends in AI-related pharmaceutical patents over the past decade. This dual approach allows for both the nuanced understanding provided by qualitative legal analysis and the empirical validation offered by quantitative data trends, creating a robust foundation for assessing how IP systems are adapting (or failing to adapt) to AI innovations in the pharmaceutical sector.

Data Sources[4,5]

The research draws upon a wide range of authoritative sources to ensure comprehensive coverage of this multidisciplinary topic. Academic literature from peer-reviewed journals in law, computer science, and pharmaceutical sciences provides the theoretical foundation, while primary legal sources such as the EU Copyright Directive (2001/29/EC) and recent amendments regarding text and data mining exceptions offer crucial regulatory context. Industry reports from organizations like the Pharmaceutical Research and Manufacturers of America (PhRMA) and the World Health Organization (WHO) supply valuable data on AI adoption rates and implementation challenges across the pharmaceutical industry. Additionally, the study incorporates white papers from leading AI research institutions and position statements from regulatory bodies like the FDA and EMA, which are increasingly addressing AI applications in drug development. These diverse sources enable a 360-degree perspective on the issue, capturing insights from legal, technological, and industry-specific viewpoints.

Limitations

Several important limitations must be acknowledged in this research. The most significant challenge stems from the jurisdictional variability in IP laws, where differences in copyright regimes, patent eligibility criteria, and fair dealing exceptions between countries complicate efforts to develop universal recommendations. For instance, what qualifies as fair use of copyrighted material for AI training in the United States may constitute infringement in certain European jurisdictions with more restrictive text and data mining exceptions. Another major limitation arises from the extraordinarily rapid evolution of AI technologies, which consistently outpaces the much slower process of legal and regulatory reform. Findings about current legal interpretations may become outdated as new AI capabilities emerge or as courts issue rulings on pending cases involving AI-generated inventions. Additionally, the proprietary nature of many commercial AI systems in pharmaceuticals limits access to complete information about their development processes and training data sources, potentially leaving gaps in understanding the full IP implications. These limitations suggest the need for ongoing research to track developments in both technology and law[6,7].

Case Studies

Pfizer's Use of AI in Drug Discovery

Pfizer, one of the world's leading biopharmaceutical companies, has integrated generative artificial intelligence (AI) and text and data mining (TDM) techniques into its drug discovery pipeline to accelerate the identification of novel therapeutic targets. In its published case study, Pfizer details how AI algorithms were deployed to analyze vast and diverse datasets, including genomic sequences, clinical trial records, biomedical literature, and proprietary research findings. By applying TDM, the company extracted relevant patterns, correlations, and mechanistic insights that would be difficult or time-consuming to uncover through conventional research methods. The AI-driven approach enabled Pfizer to rapidly filter through millions of data points to identify promising biological pathways and molecular targets with potential for therapeutic intervention. For example, machine learning models were trained to recognize disease-specific biomarkers and predict their role in disease progression, thereby helping researchers prioritize the most viable candidates for drug development. This process not only reduced the time required for early-stage research but also increased the accuracy of target identification by minimizing false positives and uncovering previously overlooked relationships[8].

Furthermore, the case study highlights how generative AI facilitated hypothesis generation and molecular design by simulating the interaction between candidate compounds and target proteins. These predictive models allowed researchers to explore a broader chemical space without the need for exhaustive laboratory synthesis and testing. By combining AI-generated predictions with expert validation, Pfizer was able

to focus resources on the most promising leads, thereby enhancing efficiency and reducing costs in the research pipeline. From an operational standpoint, Pfizer's implementation of AI in drug discovery demonstrates how TDM can be harnessed in a legally compliant manner when supported by licensed datasets and robust data governance practices. The company's approach underscores the importance of balancing innovation with ethical and regulatory considerations, particularly in managing sensitive patient data and respecting intellectual property rights associated with mined content. Overall, this case exemplifies the transformative potential of generative AI and TDM in revolutionizing pharmaceutical research while maintaining adherence to scientific and legal standards[9].

GSK's Use of AI in Medical Image Analysis

GlaxoSmithKline (GSK), a global leader in pharmaceuticals and healthcare, has explored the use of generative artificial intelligence (AI) and text and data mining (TDM) techniques in medical image analysis to improve patient care and clinical decision-making. In its published case study, GSK describes how AI systems were developed to process large volumes of medical images such as MRI scans, CT scans, and histopathology slides alongside patient health records. Using TDM, the company extracted complex image features and correlated them with clinical outcomes, enabling a more precise understanding of disease progression. By training AI algorithms on diverse, high-quality image datasets, GSK aimed to identify subtle visual patterns that may not be easily detectable by human specialists. These models were capable of segmenting tissues, detecting anomalies, and quantifying disease markers with high accuracy. For example, in oncology research, the AI could measure tumor size and morphological changes over time, while in neurology, it could assess brain structure alterations linked to degenerative conditions. Integrating these insights with other clinical and genomic data allowed the system to generate predictive models of patient outcomes, which could guide personalized treatment strategies[10].

The generative AI models also contributed to risk stratification by predicting which patients were more likely to experience disease recurrence or respond to specific treatments. This predictive capability reduced the trial-and-error approach in clinical decision-making and supported earlier interventions for high-risk individuals. Additionally, the use of synthetic data generation techniques helped GSK address limitations in data availability and privacy by creating realistic yet anonymized medical images for algorithm training and validation. From an implementation perspective, GSK's approach illustrates the value of combining generative AI with TDM to leverage both visual and textual health data. The case study also emphasizes the importance of robust validation against real-world clinical results to ensure algorithm reliability and generalizability across diverse patient populations. By embedding AI-driven image analysis into its research and clinical workflows, GSK demonstrated how advanced analytics can enhance diagnostic accuracy, optimize treatment planning, and ultimately improve patient outcomes while maintaining strict compliance with medical data governance standards.

Novartis's Use of AI in Molecular Design

Novartis, a global pharmaceutical innovator, has integrated generative artificial intelligence (AI) and text and data mining (TDM) into its molecular design processes to accelerate early-stage drug discovery. In its published case study, the company outlines how AI algorithms were employed to create entirely new molecular structures with the potential to become drug candidates. Leveraging vast datasets of chemical compounds, biological assay results, and clinical trial records, Novartis trained generative models such as variational autoencoders (VAEs) and generative adversarial networks (GANs) to propose novel molecules that matched specific pharmacological profiles. The AI models were able to virtually "design" drug-like compounds by predicting how molecular structures would interact with biological targets, such as enzymes, receptors, or proteins implicated in disease. This predictive capability extended beyond simple binding affinity, incorporating key safety and efficacy parameters into the design process. By simulating drug-target interactions and pharmacokinetic behavior computationally, Novartis significantly reduced the reliance on costly and time-consuming wet-lab synthesis and early-stage testing[11,12].

In addition to generation, the AI system employed TDM to extract and integrate knowledge from patents, research publications, and proprietary internal data. This allowed researchers to avoid pursuing molecular designs with a high risk of toxicity or redundancy with existing compounds. The models also ranked molecules by their probability of success in later stages of development, enabling more strategic allocation of resources. The use of generative AI in molecular design also provided Novartis with the flexibility to explore chemical spaces that would have been difficult to investigate through traditional medicinal chemistry approaches. Synthetic feasibility checks were built into the AI workflow to ensure that proposed molecules could be practically produced in a laboratory setting. Moreover, the ability to generate synthetic molecular data helped overcome challenges of data scarcity in niche therapeutic areas while protecting proprietary datasets from external exposure. By embedding these AI-driven molecular design capabilities into its R&D pipeline, Novartis shortened the drug discovery timeline, increased hit rates for viable drug candidates, and enhanced its

capacity to design safer, more effective therapeutics. The case study underscores the transformative potential of combining generative AI with TDM to create a data-rich, innovation-driven approach to pharmaceutical research.

Sanofi's Use of AI for Pattern Recognition in Drug Discovery and Development

Sanofi, a leading multinational biopharmaceutical company, has harnessed artificial intelligence (AI) and text and data mining (TDM) technologies to enhance its drug discovery and development workflows. In its published case study, the company highlights how AI-driven analytics were used to process and interpret massive volumes of heterogeneous biomedical data, ranging from genomic sequences and chemical compound libraries to clinical trial outcomes and real-world patient records. The primary objective was to uncover hidden patterns and complex relationships within these datasets that could guide the selection of promising drug candidates and therapeutic targets. By applying advanced machine learning algorithms including deep neural networks and graph-based models Sanofi was able to map connections between molecular structures, disease mechanisms, and patient responses. This allowed researchers to pinpoint biomarkers, predict disease progression, and identify novel intervention points that traditional statistical methods might have overlooked[13,14].

Sanofi's AI systems were also integrated with TDM pipelines capable of extracting and synthesizing insights from millions of research articles, patents, and regulatory submissions. This ensured that discoveries were contextualized within the broader scientific and regulatory landscape, reducing duplication of effort and avoiding the pursuit of drug candidates with known limitations. Furthermore, TDM accelerated competitive intelligence gathering, enabling Sanofi to adapt its R&D strategy to emerging trends and gaps in the market. The integration of AI into Sanofi's R&D framework not only streamlined hypothesis generation but also improved candidate prioritization by estimating the probability of technical and clinical success. By identifying subtle relationships such as the effect of specific genetic mutations on drug response Sanofi enhanced its capacity for personalized medicine, targeting therapies to patient subgroups most likely to benefit.

Overall, Sanofi's approach demonstrates the transformative role of AI and TDM in making data-driven decisions, shortening the discovery timeline, and increasing the efficiency of pharmaceutical innovation. The case study illustrates how a systematic application of AI for pattern recognition can lead to breakthroughs in both preclinical and clinical stages of drug development.

These case studies highlight the dual nature of AI in innovation while it offers breakthroughs like Novartis's molecular designs, it also forces a reevaluation of IP laws in the age of machine learning. Future policies must balance innovation incentives with ethical data usage to avoid stifling progress or enabling infringement.

Recommendations

Policy Reforms

To address the intellectual property challenges posed by AI in pharmaceutical research, policymakers should prioritize two key reforms. First, existing fair dealing exceptions should be explicitly expanded to include AI-driven text and data mining activities for research purposes, creating a safe harbor that protects legitimate scientific uses of copyrighted materials while respecting creator rights. This expansion should be carefully crafted to balance innovation with IP protection, potentially following the model of Japan's 2018 copyright reforms which created specific exceptions for data mining. Second, governments must develop AI-specific IP frameworks that clarify ownership of AI-generated outputs, particularly in the pharmaceutical context where research builds upon existing knowledge. A potential solution could involve shared rights models where both AI developers and data providers maintain certain rights, similar to joint authorship provisions in copyright law but adapted for machine-generated content. These frameworks should also address the patentability of AI-discovered compounds, potentially creating a new category of protection that recognizes the unique nature of machine-assisted inventions while maintaining incentives for pharmaceutical innovation[15].

Regulatory Actions

Regulatory agencies must take proactive steps to create a coherent global framework for AI in drug development. The FDA and EMA should work with the World Intellectual Property Organization (WIPO) to harmonize standards for evaluating AI-generated evidence in regulatory submissions, ensuring consistent approaches to algorithm validation and data quality assessment across jurisdictions. This harmonization should extend to IP protections, creating predictable rules for AI-assisted inventions worldwide. Additionally, regulatory bodies should mandate rigorous ethical standards for pharmaceutical AI systems, including requirements for regular bias audits to detect and correct demographic or data source biases that could affect drug safety or efficacy predictions. Explainability requirements should be implemented, compelling developers to document and disclose sufficient information about AI decision-making processes to enable proper scientific

evaluation, while protecting legitimate trade secrets through graduated disclosure mechanisms. These measures would build trust in AI applications while maintaining scientific rigor in drug development.

Industry Practices

Pharmaceutical companies and AI developers should adopt several best practices to navigate the IP landscape responsibly. Collaborative licensing models, such as data pools or patent commons specifically for AI training data, could help resolve copyright issues while accelerating research. Initiatives like the European Health Data Space could serve as models for secure, ethical data sharing in pharmaceutical AI development. For patient data used in AI training, companies must implement robust anonymization protocols that go beyond basic de-identification to include modern techniques like differential privacy, particularly when working with sensitive health information. The industry should also establish clear governance frameworks for AI development projects, defining IP rights and data usage terms at the outset of collaborations between pharmaceutical firms, AI startups, and academic institutions. These practices would not only ensure legal compliance but also foster public trust in AI-driven pharmaceutical innovations, which is essential for widespread adoption of these transformative technologies.

CONCLUSION

The integration of generative AI into pharmaceutical research marks a paradigm shift in drug discovery, clinical development, and personalized medicine, offering unprecedented speed and innovation but exposing significant gaps in intellectual property (IP) frameworks designed for human-generated work. Current laws struggle to address ownership of AI-derived discoveries, fair use of copyrighted training data, and cross-border regulatory consistency, risking both overregulation that stifles progress and legal uncertainty that deters investment. Policymakers must modernize IP protections to reflect AI's unique capabilities, enforce ethical standards for data use and transparency, and promote global harmonization through bodies like WIPO, WHO, and ICH. Future research should explore regulatory strategies in emerging markets, evaluate cross-jurisdictional harmonization, and reassess shared rights and fair dealing models as AI evolves. Proactive collaboration between technologists, legal experts, regulators, and industry will be essential to ensure AI transforms medicine within a fair, transparent, and globally coherent framework that safeguards innovation and public health.

REFERENCES

1. H. Zhong, J. Chang, Z. Yang, T. Wu, P.C. Mahawaga Arachchige, C. Pathmabandu, M. Xue Copyright protection and accountability of generative ai: Attack, watermarking and attribution Companion Proceedings of the ACM Web Conference 2023 (2023), pp. 94-9
2. Lucchi N, ChatGPT: A case study on copyright challenges for generative artificial intelligence systems, European Journal of Risk Regulation, 12 (2023) 1, doi:10.1017/err.2023.59.
3. N. Vyas, S.M. Kakade, B. Barak On provable copyright protection for generative models proceedings of the 40th International Conference on Machine Learning, 202, PMLR (2023), pp. 35277-35299
4. Margoni T & Kretschmer M, A Deeper Look into the EU text and data mining exceptions: Harmonisation, data ownership, and the future of technology, GRUR International, 71 (8) (20) (2022) 686.
5. Kathuria R et al., Implications of AI on the Indian economy, Indian Council for Research on International Economic Relations (ICRIER), 2020; To build an Encompassing Data Strategy is one of the five key recommendations of ICRIER.
6. US Chamber of Commerce, Art of the impossible, US Chamber of international IP indexglobal innovation policy centre, 2020, 67.
7. Margoni T, A deeper look into the EU text and data mining exceptions: Harmonisation, data ownership, and the future of technology, GRUR International, 71 (8) (2022) 687.
8. OECD, AI-language models: Technological, socio-economic and policy considerations, OECD Digital Economy Paper, April 2023, No. 352, 32 & 38.
9. Carroll M W, Copyright and the progress of science: Why text and data mining is lawful,University of California, Davis, 53 (893) 956.
10. Justice Singh P M, Evolution of Copyright Law: The Indian journey, Indian Journal of Law and Technology, 16 (2) (2020) 54.
11. H. Von Foerster On Self-Organizing Systems and Their Environments Understanding understanding, Springer (2003), 10.1007/0-387-21722-3_1
12. Kathuria R, Kedia M & Kapilavai S, Implications of AI on the Indian economy, Indian Council for Research on International Economic Relations (ICRIER), 2020.

13. Dermawan A & Mezei P, Artificial intelligence and consensus-based remuneration regime in Southeast Asia, SSRN, https://ssrn.com/abstract=4625850 (accessed on 7 November 2023).
14. Dermawan A, Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese "nonenjoyment" purposes?, Journal of World Intellect Property, 11 (2023) 1.
15. Kathuria R, Kedia M & Kapilavai S, Implications of AI on the Indian economy, Indian Council for Research on International Economic Relations (ICRIER), 2020.