INTERNATIONAL JOURNAL FOR LEGAL RESEARCH AND ANALYSIS

Open Access, Refereed Journal Multi Disciplinary Peer Reviewed

www.ijlra.com

DISCLAIMER

EDITORIALTEAM

EDITORS

Dr. Samrat Datta

Dr. Samrat Datta Seedling School of Law and Governance, Jaipur National University, Jaipur.Dr. Samrat Datta is currently associated with Seedling School of Law and Governance, Jaipur National University, Jaipur. Dr. Datta has completed his graduation i.e., B.A.LL.B. from Law College Dehradun, Hemvati Nandan Bahuguna Garhwal University, Srinagar, Uttarakhand. He is an alumnus of KIIT University, Bhubaneswar where he pursued his post-graduation (LL.M.) in Criminal Law and subsequently completed his Ph.D. in Police Law and Information Technology from the Pacific Academy of Higher Education and Research University, Udaipur in 2020. His area of interest and research is Criminal and Police Law. Dr. Datta has a teaching experience of 7 years in various law schools across North India and has held administrative positions like Academic Coordinator, Centre Superintendent for Examinations, Deputy Controller of Examinations, Member of the Proctorial Board

Dr. Namita Jain

Head & Associate Professor

School of Law, JECRC University, Jaipur Ph.D. (Commercial Law) LL.M., UGC -NET Post Graduation Diploma in Taxation law and Practice, Bachelor of Commerce.

Teaching Experience: 12 years, AWARDS AND RECOGNITION of Dr. Namita Jain are - ICF Global Excellence Award 2020 in the category of educationalist by I Can Foundation, India.India Women Empowerment Award in the category of "Emerging Excellence in Academics by Prime Time &Utkrisht Bharat Foundation, New Delhi.(2020). Conferred in FL Book of Top 21 Record Holders in the category of education by Fashion Lifestyle Magazine, New Delhi. (2020).Certificate of Appreciation for organizing and managing the Professional Development Training Program on IPR in Collaboration with Trade Innovations Services, Jaipur on March 14th, 2019

Mrs.S.Kalpana

Assistant professor of Law

Mrs.S.Kalpana, presently Assistant professor of Law, VelTech Rangarajan Dr.Sagunthala R & D Institute of Science and Technology, Avadi.Formerly Assistant professor of Law,Vels University in the year 2019 to 2020, Worked as Guest Faculty, Chennai Dr.Ambedkar Law College, Pudupakkam. Published one book. Published 8Articles in various reputed Law Journals. Conducted 1Moot court competition and participated in nearly 80 National and International seminars and webinars conducted on various subjects of Law. Did ML in Criminal Law and Criminal Justice Administration.10 paper presentations in various National and International seminars. Attended more than 10 FDP programs. Ph.D. in Law pursuing.

Avinash Kumar

Avinash Kumar has completed his Ph.D. in International Investment Law from the Dept. of Law & Governance, Central University of South Bihar. His research work is on "International Investment Agreement and State's right to regulate Foreign Investment." He qualified UGC-NET and has been selected for the prestigious ICSSR Doctoral Fellowship.He is an alumnus of the Faculty of Law, University of Delhi. Formerly he has been elected as Students Union President of Law Centre-1, University of Delhi.Moreover, he completed his LL.M. from the University of Delhi (2014-16), dissertation on "Cross-border Merger & Acquisition"; LL.B. from the University of Delhi (2011-14), and B.A. (Hons.) from Maharaja Agrasen College, University of Delhi. He has also obtained P.G. Diploma in IPR from the Indian Society of International Law, New Delhi.He has qualified UGC – NET examination and has been awarded ICSSR – Doctoral Fellowship. He has published six-plus articles and presented 9 plus papers in national and international seminars/conferences. He participated in several workshops on research methodology and teaching and learning.

ABOUT US

INTERNATIONAL JOURNAL FOR LEGAL RESEARCH & ANLAYSIS ISSN

2582-6433 is an Online Journal is Monthly, Peer Review, Academic Journal, Published online, that seeks to provide an interactive platform for the publication of Short Articles, Long Articles, Book Review, Case Comments, Research Papers, Essay in the field of Law & Multidisciplinary issue. Our aim is to upgrade the level of interaction and discourse about contemporary issues of law. We are eager to become a highly cited academic publication, through quality contributions from students, academics, professionals from the industry, the bar and the bench. INTERNATIONAL JOURNAL FOR LEGAL RESEARCH & ANALYSIS ISSN 2582-6433 welcomes contributions from all legal branches, as long as the work is original, unpublished and is in consonance with the submission guidelines.

TEXT AND DATA MINING: COMPARITIVE ANALYSIS OF FAIR USE AND FAIR DEALING

AUTHORED BY - RITWIK KARMAKAR

Research Scholar, CHRIST (Deemed to be University)

INTRODUCTION

Intellectual Property Rights plays a vital role in protecting the rights of creators and inventors. It includes copyright, patents, trademark, geographical indications, designs, biological diversity, integrated circuits and so on. Basically, the original work or invention is protected under this law. The Copyrights Act, 1957 was formulated in order to protect the rights of a person who claims to be the first and true owner of his work. It simply protects the right of creators so that the originality of work is preserved. This includes literary, artistic, musical, dramatic, sound recording and cinematography. The primary purpose of copyright is to give authors and creators control over the use of their works, ensuring they can benefit from their creations and prevent unauthorized exploitation.

In many jurisdictions, copyright protection is automatic once a work is created and fixed in a tangible medium, such as a written document or digital file. The rights typically granted under copyright include the right to reproduce, distribute, perform, display, and create derivative works based on the original creation. However, copyright isn't absolute and there are limitations to the rights it provides. One important limitation is the doctrine of "fair dealing" (or "fair use" in some jurisdictions). This doctrine allows for the limited use of copyrighted material without permission from the copyright holder, under specific circumstances. The doctrine of fair dealing under the Indian Copyright Act, 1957, allows for the limited use of copyrighted material without requiring permission from the copyright owner, providing a balance between protecting creators' rights and serving public interests. Section 52 of the Act specifies various exceptions, including uses for private study, criticism, review, news reporting, and educational purposes. This doctrine ensures that essential activities such as academic research and public commentary can proceed without undue restrictions, while also considering factors like the purpose of use, the nature of the work, the amount used, and its potential market impact. However, it does not specifically address text and data mining (TDM), a modern process involving the extraction and analysis of large datasets to uncover patterns or insights.

TDM often requires access to significant portions of copyrighted material, which can be challenging under current fair dealing provisions that may not fully accommodate the extensive data use inherent in these activities. Text and data mining involves the extraction and analysis of large datasets to identify patterns, trends, and relationships. This process is increasingly important in various fields, including academic research, machine learning, artificial intelligence, and business analytics. TDM relies on accessing and analyzing extensive amounts of text and data, often raising questions about copyright infringement when copyrighted materials are used. Traditionally, copyright law does not explicitly address TDM, which creates uncertainty for researchers and businesses that wish to engage in this practice. The core issue is that TDM typically requires access to copyrighted materials, which may be subject to restrictive access controls or licensing agreements. As a result, researchers and organizations may face legal challenges when using copyrighted content for TDM purposes. To understand how the Indian judiciary understands the concept of Fair Dealing, let us see the case of *Kartar Singh Giani v. Ladha Singh*,[1] whereby the Hon'ble Bench observed that:

"Two points have been urged in connection with the meaning of the expression fair, in fair dealing. First, that in order to constitute unfairness there must be an intention to compete and to derive profit from such competition and secondly, that unless the motive of the infringer were unfair in the sense of being improper the dealing would be fair."

The Doctrine of Fair Dealing was initially sparked with an understanding of the doctrine of equity which allows certain secondary use of certain copyrighted work, which would amount to a clear infringement of the copyright. In other words, if the secondary use of the copyrighted product is fair and just, the secondary use would rather fall under the ambit of the exception instead of the ambit of copyright infringement.[2]

FAIR DEALING IN DIGITAL AGE

With the presence and increasing use of the Internet the wholly new global world has emerged, and the protection of copyrighted work is the major concern. Legislation across the globe have incorporated technological protection measures in their copyright regimes, so as to better protect or rather restrict, the access and use of copyrighted work.[3] These technologies like much

---

[1] Kartar Singh Giani v. Ladha Singh (1935) 16 ILR LAH 103

[2] Vaibhavi Pandey, India: "Fair Dealing" In Copyrights: Is The Indian Law Competent Enough to Meet the Current Challenges?MONDAQ(Oct. 24, 2019, 2:00PM) http://www.mondaq.com/india/x/299252/Copyright/Fair+Dealing+In+Copyrights+Is+The+Indian+Law+Compe tent+Enough+To+Meet+The+Current+Challenges.

[3] Megha Nagpal, Copyright Protection through digital Rights Management in India: A Non-Essential Imposition, 22 JOURNAL OF INTELLECTUAL PROPERTY RIGHTS 224, 227-230 (2017).

innovation are both promising and potentially harmful to various parties interested in the use and exploitation of work of authorship- from book and music to film and web pages.

The nature of computer programs is difficult to determine with the idea/expression distinction. The computer programs are utilitarian articles that accomplish a task which can me logical, structural, etc. which are dictated by the functions in the programs with certain external inputs.[4] But, because of the hybrid nature of computer programs, there cannot be a set of rules or factors that can be used in identifying the unprotected idea and the protected expression. The Third Circuit attempted to resolve the idea/express distinction of computer programs by observing that the "*idea or function of a computer programs is the idea in itself as a whole and everything else which is not necessary to that purpose or function could be construed as the expression of that idea.*"[5] Let us understand how the United States of America have used the exception of fair use of copyrighted work in the digital age. Let us try and understand the shift of revenue of the music labels from selling Vinyl Records to Digital downloads to Online streaming. In the following chart, the International Federation of Phonographic Industry (hereby referred to as "**IFPI**") conducted research on the revenue earned by the music labels from 1999 to 2017.[6]

In the above graph, we see that the primary source of the revenue of the music label organization shifted from the physical sale to the digitized sale. The revenue started dropping immensely with the introduction of the digital P2P file sharing and downloads which can been as the orange line in the above graph. Further, with the introduction of the online music streaming platforms like Spotify, apple music, Sound Cloud and YouTube from 2006, the revenues again shifted from digital downloads of sound recordings to online streaming platform which can be seen in the grey line in the above graph. To further prove the point, the following graph will show another research by an online website called Statista regarding how online streaming platforms came as the messiah for the music label from the apocalypse of music industry by the demon called piracy.[7]

---

[4] Computer Association Intern., Inc. v. Altai, 982 F.2d 693 (2nd Cir. 1992).

[5] Whelan Associates v. Jaslow Dental Laboratory, 797 F.2d 1222 (3d Cir. 1986).

[6] International Federation of the Phonographic Industry, Global Music Report 2018: State of the Industry, INTERNATIONAL FEDERATION OF THE PHONOGRAPHIC INDUSTRY (OCT. 24, 2019, 6:30 PM) https://www.ifpi.org/downloads/GMR2018.pdf.

[7] Id.

TEXT AND DATA MINING

In today's world, the amount of data being generated is enormous, from social media interactions to online transactions. However, this data is often in raw, unstructured forms such as text, images, or numbers, making it hard to extract meaningful insights. This is where Text and Data Mining (TDM) comes in. TDM involves the use of various computational techniques to analyze large datasets and uncover hidden patterns, relationships, and trends. The goal is to transform this raw data into useful knowledge that can be applied to solve problems or innovate in various fields. As data grows in volume, variety, and complexity, the importance of TDM in research, business, healthcare, and technology continues to rise. TDM helps unlock the value in both structured data (like spreadsheets) and unstructured data (like social media posts or research papers), making it a powerful tool in many different sectors.

Text and Data Mining is essentially a way of analyzing data to uncover patterns or information that aren't immediately obvious. It involves two main components:Text Mining (TM): This focuses on extracting useful information from text-based data, like research papers, social media posts, or customer reviews. Techniques such as Natural Language Processing (NLP) are used to analyze human language and identify trends, sentiments, or key topics within the text. For instance, sentiment analysis can tell whether people are happy, sad, or angry based on the language they use in social media posts. Data Mining (DM): This focuses on identifying patterns in structured data, such as numbers or categories, using algorithms and statistical methods. This technique is commonly used to predict future trends or behaviors. For example, it might be used to predict which customers are most likely to buy a product based on their previous purchases. Together, text mining and data mining help organizations process and make sense of the massive amounts of data available to them, turning raw data into actionable insights.

THE IMPORTANCE OF TDM IN RESEARCH AND INNOVATION

TDM is not just a tool for business; it plays a crucial role in advancing research and innovation. Whether in academic research, medical breakthroughs, or business decisions, TDM helps researchers and organizations gain a deeper understanding of complex problems and discover new opportunities.

1. TDM in Scientific Research

   One of the primaries uses of TDM is in scientific research. With the explosion of

academic publications in recent years, researchers often struggle to keep up with the sheer volume of new information being published. TDM allows researchers to mine academic papers, books, and journals quickly to identify key trends, discover new research topics, or track the evolution of particular fields. For instance, in medicine, researchers use TDM to analyze clinical trial data, medical records, and research papers to find patterns that could lead to new treatments. This helps researchers make discoveries faster than ever before, potentially saving lives.

2. TDM in Business and Market Analysis

TDM is also revolutionizing the business world. Companies use TDM techniques to understand customer behaviors, identify market trends, and improve their products or services. For example, online retailers like Amazon analyze customer reviews and feedback to identify which products are popular and which might need improvements. Additionally, TDM can help businesses understand how customers feel about their brand or products. Sentiment analysis of social media posts, product reviews, and online comments can help businesses improve customer satisfaction by addressing negative feedback quickly and adapting to market preferences.

3. TDM in Healthcare

In healthcare, TDM is transforming patient care and medical research. By analyzing electronic health records (EHR), doctors can identify at-risk patients, predict disease outbreaks, and even detect early signs of diseases like cancer or heart disease. Researchers use TDM to analyze medical literature, clinical trials, and patient data to find new treatments and drug interactions that can improve patient outcomes. For example, in cancer research, TDM helps identify patterns in genetic data, potentially leading to breakthroughs in cancer treatment. By mining these large datasets, researchers can find insights that may not be obvious from just looking at a few individual cases.

APPLICATION OF TDM ACROSS SECTORS

TDM is used in various industries and sectors to solve problems, improve services, and drive innovation. Below are some key examples of how TDM is applied:

1. Academic Research

TDM is a great tool for researchers who need to sift through thousands of academic papers. It can be used to track research trends, analyze citations, and even identify potential research gaps. Researchers in fields like climate change or physics use TDM

to analyze vast datasets and uncover new trends or findings that might otherwise be missed.

2. Social Media and Sentiment Analysis.

    Social media platforms generate enormous amounts of data every day. Companies and researchers use TDM to analyze social media posts to understand public opinion. Sentiment analysis can help businesses understand how people feel about a product, service, or political event based on their online comments. By monitoring social media in real-time, companies can make quick decisions to address customer concerns or capitalize on positive trends.

3. Fraud Detection in Financial Services

    TDM is widely used in financial services to detect fraud and monitor transactions. By analyzing patterns in transaction data, banks can spot unusual behavior that might suggest fraudulent activity, such as a sudden change in spending patterns or transactions from unfamiliar locations. TDM helps financial institutions quickly identify potential fraud before it becomes a serious issue.

4. Artificial Intelligence and Machine Learning

    Text and data mining also play a crucial role in developing artificial intelligence (AI) and machine learning (ML) models. AI systems like voice assistants (Siri, Alexa) or chatbots rely on TDM to understand human language and respond accordingly. Machine learning models also use TDM techniques to train on large datasets, allowing them to make predictions or decisions based on historical patterns. For example, AI can predict customer behavior or recommend products based on past interactions.

5. E-commerce and Personalization

    In the e-commerce world, TDM is used to improve customer experiences and increase sales. Online retailers use TDM to analyze past purchasing behavior and recommend products that a customer is likely to buy next. This helps create a more personalized shopping experience, leading to higher customer satisfaction and increased sales. For example, Amazon uses TDM to recommend products based on what a customer has viewed or purchased in the past, while Netflix uses TDM to suggest movies and TV shows based on viewing history.

CHALLENGES OF TEXT AND DATA MINING

While **Text and Data Mining (TDM)** has many benefits, it also faces several challenges that can make it difficult to implement and use effectively. These challenges include issues related

to data privacy, data quality, and the complexity of the techniques involved. Understanding these challenges is important for anyone who wants to use TDM in their work, whether in research, business, or healthcare. **Data Privacy and Security.** One of the biggest challenges in TDM is ensuring **data privacy** and **security**. When mining data, especially personal data, there are serious concerns about how this data is collected, stored, and used. In many cases, TDM involves analyzing large datasets that may contain sensitive information, such as customer details, medical records, or financial data. For example, companies that use TDM to analyze customer behavior need to make sure they are not violating privacy laws, such as the **General Data Protection Regulation (GDPR)** in Europe. GDPR sets strict rules on how personal data should be handled, ensuring that individuals' data is protected. Organizations must be careful to anonymize data, get proper consent, and make sure that their methods comply with privacy regulations. Without this, there could be legal issues, loss of trust, and reputational damage.

Another major challenge is the **quality of the data** being analyzed. For TDM to work effectively, the data must be clean, complete, and accurate. Unfortunately, many datasets are messy, with missing values, errors, or inconsistencies. This can make it difficult to draw accurate conclusions from the analysis. For example, in healthcare, if medical records are incomplete or have mistaken, TDM could lead to wrong predictions about a patient's health. Similarly, if a company is using social media data for sentiment analysis, poorly written posts, slang, or misspellings might make it harder to analyze the data accurately. Improving data quality often requires significant effort in cleaning and preparing the data before it can be analyzed. This process can be time-consuming and costly, but it is essential to ensure that the results of TDM are reliable and trustworthy. TDM involves using advanced algorithms, statistical methods, and machine learning models to analyze large datasets. These techniques can be very complex, and it often requires specialized knowledge to use them correctly.

The algorithms used in TDM are not always easy to understand, and choosing the right approach for a particular type of data or problem can be difficult. For example, text mining, which involves analyzing text data, requires techniques such as **Natural Language Processing (NLP)**. NLP helps computers understand and process human language, but it is a challenging task, especially when dealing with languages that have many nuances, slang, or informal terms. Similarly, when analyzing data from different sources, it may be hard to choose the right data mining technique, as different types of data (such as numerical, textual, or image data) require

different approaches. To overcome this challenge, it's important to have experts in data science and machine learning who understand the techniques involved and can apply them effectively. Unfortunately, not all organizations have access to such specialized talent, making TDM more difficult for some to implement. Once TDM techniques have been applied to a dataset, interpreting the results can be challenging. While TDM can find patterns and trends in data, these findings may not always be easy to understand or actionable. For example, a company may use TDM to find patterns in customer feedback, but the results may be too complex or vague to lead to clear business decisions. In many cases, TDM results need to be explained and contextualized by experts to make sense of them. This requires both technical expertise in data science and a deep understanding of the subject matter (e.g., healthcare, business, or social science) to ensure that the findings are useful and meaningful.

OVERVIEW OF SECTION 52

Section 52 of the Indian Copyright Act explicitly lists several actions that do not constitute copyright infringement. These actions, termed as **exceptions** to infringement, allow users to access and use copyrighted works without seeking permission from the copyright holders, under certain conditions. These exceptions are designed to serve public policy objectives, including **education, research, and the free flow of information.** For **Text and Data Mining,** this section is particularly important because TDM often involves the use of **large amounts of copyrighted text and data** to extract patterns, insights, or build models. By considering these practices under specific exceptions, Section 52 helps foster an environment conducive to innovation and academic research. Section 52 of the Indian Copyright Act, 1957 provides several important exceptions that allow the use of copyrighted works without infringement. These exceptions are significant in the context of Text and Data Mining (TDM), especially when the use is for research or educational purposes. Subsection 1(a) allows for the reproduction of a copyrighted work for private and personal use without being considered infringement. This exception permits individuals to copy works for personal, non-commercial purposes, such as studying or creating a personal archive. However, it is not directly applicable to TDM, as TDM generally involves the large-scale use of texts, which extends beyond personal, non-commercial use.

This provision directly supports **academic research** and **private study,** permitting the use of copyrighted works for analysis, study, or research without requiring permission from the

copyright holder. TDM activities conducted for research purposes are likely covered by this exception, as researchers often use large datasets (including text and data from books, articles, and websites) for non-commercial, academic work. **Text and Data Mining** (TDM) carried out for **research purposes,** such as academic studies or data analysis, falls under this exception. If the data used is purely for **academic research** or non-commercial purposes, the mining of copyrighted content might be permitted under this exception.

IS TEXT AND DATA MINING COVERED UNDER SECTION 52?

Text and Data Mining (TDM) involves extracting useful information and patterns from large datasets, often containing copyrighted material such as books, articles, or websites. TDM typically requires significant amounts of text data to be accessed, reproduced, and processed. **Section 52's** focus on **research, educational purposes, and non-commercial use** makes it relevant for TDM, especially in the following scenarios:

- **Non-Commercial Research:** If TDM is conducted for academic or research purposes, and not for commercial profit, it could potentially fall under the exception for research and study outlined in **Section 52(1)(b).**

- **Use in Educational Institutions:** If TDM is carried out within an educational institution for teaching, research, or other academic purposes (as covered in **Section 52(1)(i)),** this would also be an allowable exception.

- **Libraries and Archives**: Research projects conducted by libraries or archives, particularly those involving TDM to analyze or digitize text for preservation or study, may be allowed under the non-commercial provisions of **Section 52(1)(c).**

However, TDM does not automatically fall under these exceptions. It depends on the context of the mining activities, the **nature of the data** used (whether it is copyrighted material), and whether the activities are **commercial** or **non-commercial.**

Issues with Text and Data Mining under Section 52

While **Section 52** provides several exceptions for research and educational use, there are **legal and practical challenges:**

- **Scale of Use:** TDM often involves large-scale extraction and processing of data, which may extend beyond what is typically considered **private study** or **research**. The scale of the activity may push it into the realm of commercial use, which would not be protected under Section 52 exceptions.

- **Type of Data**: TDM typically uses **digitized text** and **online content**. If the text is **not sufficiently transformed** or is used in ways that go beyond the scope of research, it may not be considered "fair use" under Section 52.
- **Commercial Use**: If TDM is conducted with commercial objectives (e.g., to build a proprietary database for resale or to create a commercially sold product), it may not qualify for the research or education exceptions, making it potentially subject to copyright infringement.

CHALLENGES OF APPLYING THE COPYRIGHT ACT TO TEXT AND DATA MINING (TDM)

Text and Data Mining (TDM) refers to computational techniques used to extract patterns, trends, and insights from large datasets. TDM is used across many fields like natural language processing (NLP), artificial intelligence (AI), bioinformatics, and more. The application of TDM often requires researchers to work with significant amounts of data, much of which is copyrighted. However, the **Indian Copyright Act of 1957** presents several challenges when it comes to reconciling the need for TDM with copyright protection. The Indian Copyright Act provides certain exclusive rights to authors, including the right to **reproduce, distribute,** and **adapt** their works. These rights are meant to protect the economic interests of creators and ensure that their works are not used without permission.

Reproduction of Copyrighted Material: TDM tools typically involve copying large portions of text, data, or other media from copyrighted works to analyze them. Under the Indian Copyright Act, copying copyrighted works for any purpose beyond personal use or research may be considered **infringement.** The issue arises because, in order to perform TDM, significant chunks of copyrighted content must be copied and analyzed. Even if the ultimate goal is not to reproduce the content for commercial purposes, this process may still fall under the reproduction right and thus infringe the copyright.

Adaptation and Transformation: TDM can result in the transformation of copyrighted material. For example, researchers might use copyrighted text to generate new insights, data, or predictions. The transformation of copyrighted content into a new form is often seen as a **derivative work,** which is typically an exclusive right of the copyright owner. In India, creating a derivative work without the author's permission is an infringement. Therefore, even if the

goal of TDM is purely research-oriented and non-commercial, the transformation of copyrighted text could still be problematic from a legal standpoint.

Data Scraping and Aggregation: Many TDM methods involve **web scraping**, which is a technique for extracting data from websites. This often includes scraping publicly available text, images, and datasets. While scraping public domain content may not infringe copyright, scraping data from copyrighted sources, such as news articles, books, or academic papers, without permission can violate the copyright holder's rights. Websites often include terms of use that prohibit scraping, and breaching these terms could lead to legal consequences for researchers.

One of the major conflicts in this context is the tension between the **economic rights of copyright holders** and the **needs of researchers** engaged in digital research. Copyright law is designed to provide protection to creators, ensuring that they can control how their works are used and are compensated for their labor. However, TDM and other research methods often require access to vast amounts of data, including copyrighted materials, in order to extract useful insights. Here are the key points of conflict:

Access to Data for Research: For TDM to be effective, researchers often need access to **large datasets** that may include copyrighted works. These datasets are frequently locked behind **paywalls** or governed by restrictive **licensing agreements,** which limit access. For example, academic articles, books, and datasets hosted by publishers may only be accessible to institutions that can afford expensive subscriptions. This creates a significant barrier for researchers, particularly in developing countries like India, where access to academic content may be limited. **Copyright protection** can thus hinder the ability of researchers to gather data necessary for their work.

Open Access and Data Availability: Open access publishing and open data models allow for broader dissemination of research and provide researchers with **freely accessible materials.** However, not all copyrighted content is available through open-access licenses. For researchers involved in TDM, the lack of **open-access content** or **permissive licenses** presents a major challenge. The need for free access to data becomes even more critical in modern research areas like artificial intelligence (AI), where large-scale data analysis is essential to the development of algorithms and technologies. The conflict arises because **copyright holders**, including

publishers and authors, may be unwilling to freely provide their works for such uses, fearing a loss of control and profit.

Fair Dealing vs. Exclusive Rights: The Indian Copyright Act includes **fair dealing provisions,** which allow certain uses of copyrighted works **without authorization.** These include uses for **research, private study, criticism, or review.** However, the application of these provisions to digital research and TDM is problematic. The fair dealing exception in India is not explicitly designed to accommodate modern digital research methods like TDM. The **non-commercial nature** of fair dealing in India restricts its use in certain contexts, especially when the research has commercial applications, such as in AI development or when private companies fund research. Moreover, the **lack of clear guidelines** for TDM under fair dealing provisions leaves researchers uncertain about whether their use of copyrighted materials for TDM is legally safe.

The Issue of Licensing: Many publishers, institutions, and content providers impose restrictive **licensing models** that **limit how copyrighted works can be used.** For instance, digital content might be available for reading, but not for copying or analyzing. These restrictions conflict with the needs of TDM, where data is often required to be processed in ways that involve reproducing or adapting the original works. Researchers may not be able to obtain the necessary permissions to use the content for TDM, especially when content is locked behind exclusive **licenses or terms of service agreements** that prohibit such activities. As a result, researchers may need to engage in legally complex negotiations or face potential litigation risks.

Another significant challenge is the **legal uncertainty** surrounding TDM and copyright law. Because the Indian Copyright Act does not contain explicit provisions for TDM or digital research, researchers are often left with limited guidance on how the law applies to their activities. This lack of clarity creates a situation where researchers may be uncertain about whether their use of copyrighted works falls within the scope of **fair dealing** or **copyright infringement**. This uncertainty could discourage researchers from engaging in TDM or lead to the **risk of legal action** from copyright holders.

Risk of Litigation: The risk of **copyright infringement lawsuits** poses a significant deterrent to TDM research. If researchers are unsure about the legality of their activities, they may avoid using certain copyrighted materials, thus stifling innovation. On the other hand, if they proceed

without sufficient legal clarity, they could face legal action from copyright holders, resulting in penalties, costly settlements, or the destruction of research data.

CONCLUSION

The Indian Copyright Act of 1957 has been established to protect the rights of creators by granting exclusive control over the reproduction, distribution, and adaptation of their works. The fair dealing exceptions contained in Section 52, which allow the use of copyrighted material for purposes such as private study, research, criticism, and education, serve to balance the interests of creators with those of the public. The judicial interpretation provided in cases such as Kartar Singh Giani v. Ladha Singh reinforces that the intention behind the use of copyrighted material plays an important role in determining whether such use is permissible under the law.

Text and Data Mining (TDM) represents a significant advancement in research methodology, permitting the extraction and analysis of large volumes of text and data. This process has become indispensable in academic inquiry, commercial analysis, and technological development. TDM often requires the reproduction of substantial portions of copyrighted works to uncover patterns and insights that are not immediately apparent. When carried out for academic or non-commercial purposes, TDM may fall within the exceptions outlined in Section 52. However, the extensive nature of data extraction and the transformation involved in TDM pose practical and legal challenges. The vast scale of copying and processing large datasets raises questions regarding the extent to which existing legal provisions can accommodate modern digital research techniques.

Researchers and institutions encounter difficulties when access to digitized content is limited by strict copyright controls. The ambiguity in the law regarding the application of fair dealing to TDM creates uncertainty, which may deter scholars from utilizing this important research tool. The current framework does not clearly define the limits of acceptable use for digital research, and the risk of litigation remains a serious concern for those engaged in large-scale data analysis. This situation calls for a thorough reexamination of the legal provisions to ensure that the rights of copyright holders are maintained while supporting academic and commercial inquiry.

In summary, although Section 52 provides essential exceptions for the use of copyrighted material, its application to TDM remains uncertain and complex. A revised legal framework that clarifies these issues would enable researchers to conduct digital analyses with confidence, ensuring that creative rights are respected while fostering innovation. Policymakers must urgently address these challenges with timely and clear reforms.