The Problem of Copyright Infringement in Text and Data Mining (TDM) Under the Copyright Act B.E. 2537: A Comparative Study Between Thailand, the United States, the United Kingdom, Japan and China
ปัญหาการละเมิดลิขสิทธิ์ภายใต้พระราชบัญญัติลิขสิทธิ์ พ.ศ. 2537 ของการทำเหมือง ข้อความและข้อมูล: การศึกษาเชิงเปรียบเทียบระหว่างประเทศไทย สหรัฐอเมริกา สหราชอาณาจักร ประเทศญี่ปุ่น และประเทศจีน

Pornpepat Suwannachairob[1] and Poomsiri Dumrongvute[1]

พรพีภัทร สุวรรณชัยรบ[1] และ ภูมิศิริ ดำรงวุฒิ[1]

[1]Master of Laws in Business Law (International Program), Faculty of Law, Chulalongkorn University

[1]หลักสูตรนิติศาสตรมหาบัณฑิต สาขากฎหมายธุรกิจ (หลักสูตรนานาชาติ) คณะนิติศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

Abstract

This article examines the problem of copyright infringement in the context of Text and Data Mining (TDM) under Thailand's Copyright Act B.E. 2537 (1994), comparing it with the legal frameworks of the United States, the United Kingdom, Japan, and China. TDM, a key process in extracting insights from large datasets, often requires reproducing copyrighted materials, raising copyright infringement concerns. While countries like Japan and the U.S. have adopted flexible or specific exceptions facilitating TDM, Thailand's current legal structure, particularly Section 32 and its interpretation by the Supreme Court, offers limited scope for such activities even for non-commercial purposes. This restrictiveness hinders innovation and places Thailand at a competitive disadvantage. The study highlights the need to amend Thai copyright law to explicitly permit TDM, ensuring both legal clarity and support for technological advancement in line with global practices.

Keywords: Text and data mining (TDM); Copyright Infringement; Fair use doctrine; Copyright infringement exception

บทคัดย่อ

บทความนี้มุ่งศึกษาปัญหาการละเมิดลิขสิทธิ์จากการทำเหมืองข้อความและข้อมูลภายใต้พระราชบัญญัติลิขสิทธิ์ พ.ศ. 2537 โดยเปรียบเทียบกับกฎหมายลิขสิทธิ์ของประเทศสหรัฐอเมริกา สหราชอาณาจักร ญี่ปุ่น และจีน ทั้งนี้เนื่องจาก การการทำเหมืองข้อความและข้อมูล แม้จะเป็นกระบวนการสำคัญในการสกัดข้อมูลเชิงลึกจากชุดข้อมูลที่มีขนาดใหญ่ และซับซ้อนก็ตาม แต่กระบวนการข้างต้นจำต้องต้องอาศัยการทำซ้ำงานอันมีลิขสิทธิ์ ซึ่งเป็นการละเมิดลิขสิทธิ์ตาม

พระราชบัญญัติลิขสิทธิ์ดังกล่าว อย่างไรก็ดีประเทศต่าง ๆ เช่น ญี่ปุ่นและสหรัฐอเมริกาได้มีข้อยกเว้นการละเมิดลิขสิทธิ์ที่เอื้อต่อการทำเหมืองข้อความและข้อมูล ขณะที่กฎหมายลิขสิทธิ์ไทยโดยเฉพาะมาตรา 32 และแนวคำพิพากษาของศาลฎีกา ยังตีความอย่างจำกัด อันส่งผลให้การทำเหมืองข้อความและข้อมูลในประเทศไทยหากมีการนำงานอันมีลิขสิทธิ์มาทำซ้ำแล้ว การทำเหมือนข้อความและข้อมูลนั้นไม่สามารถทำได้ แม้เป็นการใช้เพื่อการวิจัยโดยไม่แสวงหากำไรก็ตาม ข้อจำกัดนี้ส่งผลกระทบต่อการพัฒนานวัตกรรมและทำให้ประเทศไทยเสียเปรียบในเวทีการแข่งขันระดับโลก ด้วยเหตุนี้ บทความนี้จึงเสนอให้มีการปรับปรุงข้อยกเว้นการละเมิดลิขสิทธิ์ในมาตรา 32 เพื่อให้รองรับการทำเหมืองข้อความและข้อมูล ทั้งนี้เพื่อสร้างความชัดเจนทางกฎหมายและสนับสนุนการพัฒนาเทคโนโลยีให้ทันต่อการเปลี่ยนแปลงระดับนานาชาติ

คำสำคัญ: การทำเหมืองข้อความและข้อมูล; การละเมิดลิขสิทธิ์; หลักการใช้ลิขสิทธิ์ที่เป็นธรรม; ข้อยกเว้นการละเมิดลิขสิทธิ์

Introduction

In today's digital era, the exponential growth of data—structured, semi-structured, and unstructured—has made Text and Data Mining (TDM) an essential tool for extracting insights in fields such as science, finance, and healthcare (OECD, 2013). Beyond these technical applications, TDM also holds significant potential for broader public benefit, including accelerating scientific discovery, improving healthcare outcomes, supporting evidence-based policymaking, and enabling data-driven journalism and education. TDM involves the automated copying and analysis of large volumes of content to identify patterns and trends that humans cannot easily identify through manual review (Fayyad et al., 1996; Hassani et al., 2020). However, because TDM often entails the reproduction of copyrighted works, it raises significant legal concerns under copyright law (Rosati, 2019). Specifically, since TDM requires the copying of extensive amounts of copyrighted material for analytical objectives, it may unintentionally be considered infringement under rigid copyright systems, creating legal uncertainty for innovators.

This article addresses the critical legal gap concerning TDM in Thailand by examining the inhibiting conditions for innovation that stem from the current copyright framework. Globally, various jurisdictions have addressed this issue differently. Japan has explicitly legalized TDM for both commercial and non-commercial use. The United Kingdom provides a statutory exception, albeit limited to non-commercial research. The United States relies on the flexible fair use doctrine, allowing certain TDM activities if deemed transformative. China, meanwhile, maintains a closed-list system, though its courts have occasionally accepted transformative use arguments.

In Thailand, Section 32 of the Copyright Act B.E. 2537 (1994) offers a limited exception for non-profit research, but Supreme Court interpretations have significantly narrowed its application. Even minimal commercial use or any activity that interferes with the copyright owner's ability to exploit their work may not qualify as fair use. These legal uncertainties hinder both academic and commercial TDM, placing Thailand at a disadvantage in the data-driven global economy.

This study critically examines the limitations of Thai copyright law in the context of TDM and compares it with approaches in

the United States, United Kingdom, Japan, and China. The analysis employs a doctrinal and comparative legal methodology to evaluate statutory provisions, judicial precedents, and policy directions related to TDM exceptions in these key jurisdictions. The findings are intended to inform and support the modernization of Thai copyright law, proposing a path toward enabling technological advancement while continuing to protect the rights of copyright owners.

The TDM Process and Associated Copyright Risks

While TDM is a technical process for extracting knowledge from data, its fundamental steps directly intersect with copyright law. Understanding this technical workflow is essential for identifying where and how potential copyright infringement may occur.

TDM is an advanced method that uses computer-based techniques to find useful patterns, trends, and insights hidden within large amounts of text and data (Losiewicz et al., 2000). It involves several important steps that help turn unorganized information into meaningful and practical knowledge.

1. Data Collection: The process starts by finding and accessing the right sources of information (Rosati, 2019). These sources can include structured databases or more loosely organized content like documents, websites, and articles. It's important to make sure the data is used ethically, ideally from open-access sources or with the proper permissions when required (Weber & Locke, 2022).

2. Copying of Content: After locating the relevant information, the next step is to systematically collect and duplicate it to create a "corpus"—a well-organized set of texts prepared for analysis (Natasha Karanja, 2022). This

often means pulling content from various formats such as XML files, PDFs, or web pages. These materials are saved in digital storage to support the technical processes that follow.

3. Data Preprocessing: This step prepares the data for analysis by cleaning and organizing it into a usable format. The goal is to make unstructured or messy data easier for computers to understand while keeping its essential meaning (Rosati, 2019).

4. Analysis and Mining: Now, with the cleaned data, researchers apply specific techniques to uncover useful patterns and insights based on their goals. The results are often presented using visual tools like charts, graphs, or reports (Fayyad et al., 1996). This helps to check whether the chosen methods are effective and whether the data is good enough for drawing meaningful conclusions.

As previously discussed, copyright infringement is most likely to occur during specific stages of TDM process. One such stage is data collection and access. At this point, infringement may arise if copyrighted content is accessed or collected without the necessary permissions. It is critical to distinguish between public accessibility and the legal right to reproduce content; just because material is public does not mean it is free from copyright protection, and its reproduction may still be an infringement unless a specific license or legal exception applies. Additionally, some websites have terms and conditions that prohibit automated data gathering. Accessing and using content from restricted sources, such as subscription-based databases, without a proper TDM license can also lead to potential legal issues.

Another high-risk stage is the copying of content. This is a fundamental step in TDM,

where selected materials are duplicated to form a dataset for analysis. Creating these copies, including the transient or incidental copies that may be stored temporarily during the process, may infringe on the copyright holder's exclusive rights if done without permission. Since TDM generally involves reproducing large volumes of text or data, this stage significantly increases the risk of copyright infringement. Ultimately, whether these actions constitute infringement depends on specific national copyright laws, as a country's legal framework dictates what is permissible.

Thailand's Legal Challenges to TDM under the Copyright Act B.E. 2537 (1994): A Comparative Perspective

Under Thailand's Copyright Act B.E. 2537 (1994), TDM faces major legal challenges because it involves reproducing content, an act that falls under the exclusive rights of copyright holders as outlined in Section 15. As a result, using copyrighted works in TDM without permission can be considered copyright infringement under Section 27. Whether TDM is legal in Thailand depends largely on whether it fits within the legal exceptions to infringement, particularly those found in Section 32, and on how these exceptions are interpreted by the Thai Supreme Court.

There are certain cases where TDM can be done legally. The most straightforward situation is when the copyright owner gives permission or issues a license allowing TDM. TDM is also allowed when the data being used is not protected by copyright.

This includes content that does not meet the criteria for copyright protection under Sections 4 and 6, such as non-expression of ideas, works without originality, or those not classified as protectable types of work.

It also includes works excluded from protection under Section 7, like laws, government reports, or simple facts and news.

Furthermore, works that have entered the public domain, meaning their copyright term has expired under Sections 19 to 26, can also be used freely for TDM.

On the other hand, TDM often risks being seen as copyright infringement in many common cases. If the TDM is done for research that has any commercial aspect, it is likely to be considered infringement. The Thai Supreme Court has interpreted the "not for profit" condition in Section 32 very narrowly. To qualify for an exception, an act must not only fit a specific purpose listed in Section 32's second paragraph (such as "research or study which is not for profit") but must also satisfy the two general conditions in its first paragraph. Specifically, the act must: (1) not conflict with the normal exploitation of the work by the copyright owner, and (2) not unreasonably prejudice the legitimate rights of the copyright owner. TDM's fundamental need to reproduce entire works, or substantial portions thereof, challenges these conditions directly, particularly when research outputs might serve commercial objectives. The large-scale copying required to create a TDM corpus can be interpreted as conflicting with the owner's normal exploitation and is often viewed by courts as unreasonably prejudicial to the owner's economic rights. Previous rulings, such as Decisions No. 6525/2559 and No. 5843/2543, show that even small commercial benefits or indirect profit can make this exception invalid. Even in purely non-commercial research, if TDM involves copying large parts or entire works, it may still be infringing. Section 32 requires that

the act must not interfere with the copyright holder's ability to profit from the work and must not unfairly harm their rights. Since TDM often needs to copy substantial portions of works to be effective, this becomes a legal issue. Past court decisions, including No.5469/2552 , have shown that such extensive copying is seen as harmful to the copyright owner's interests and does not qualify for protection under Section 32.

In conclusion, based on the current Thai Copyright Act and how it has been interpreted by the courts, TDM is unlikely to be protected by existing copyright exceptions. Commercial TDM is especially limited due to strict rules against profit-making. Even non-commercial TDM faces legal uncertainty, especially when it involves copying large portions of copyrighted works. This restrictive legal environment extends beyond commercial interests, affecting broader societal goals by restricting academic freedom and limiting access to knowledge, which in turn hinders Thailand's innovation ecosystem. The absence of a robust TDM exception inhibits the nation's digital development agenda and the achievement of its AI strategy, limiting the ability of local researchers to compete globally, especially in critical public-interest fields such as health, education, science, and innovation.

Comparative Legal Perspectives on TDM

1. The United States: In the United States, TDM is not specifically addressed by any dedicated law. Instead, it falls under the broader and more flexible concept of fair use, as outlined in Section 107 of the U.S. Copyright Act. Fair use allows certain uses of copyrighted content without needing permission from the copyright owner (Carroll, 2019). Whether a particular use qualifies as fair use is judged individually, based

on four key factors: the purpose and character of the use, the nature of the original work, the amount and significance of the portion used, and the impact on the market value of the original work (Samuelson, 2009).

One of the most important aspects of this test is whether the use is "transformative" —that is, whether it adds something new or has a different purpose from the original work, rather than simply copying it (Carroll, 2019). For example, in Authors Guild v. Google, Inc. (2015), the court found that Google's searchable book database, used for research and data analysis, was transformative and thus qualified as fair use. However, the limits of this principle were seen in Thomson Reuters v. ROSS Intelligence Inc. (2023), where using copyrighted legal summaries to train a competing AI system was ruled not transformative. The court also found that it harmed the market for the original content, so it did not qualify as fair use. These cases show that the legality of TDM under U.S. law depends heavily on how the specific use is carried out and its effect on the original work's value. Unlike Thailand's rigid exceptions, the U.S. fair use doctrine provides a flexible framework that has allowed courts to assess TDM through a transformative use lens, offering a pathway to legality that is absent in Thai law.

The U.S. fair use approach has several advantages. Its flexibility allows it to adapt to new technologies like TDM and artificial intelligence without needing new laws. Because fair use is developed through court decisions, it can quickly respond to innovations and remain relevant over time (Samuelson, 2015). Also, many TDM applications are naturally transformative—they aim to analyze data, find patterns, or generate insights, which is quite different from simply

republishing or reusing the original content.

However, this model also comes with serious challenges. The biggest problem is legal uncertainty. Since fair use is decided case by case in court, it's often unclear whether a specific TDM project is legal until a ruling is made. This lack of clarity can discourage innovation and investment, especially among smaller companies or research groups that may not have the resources to face potential lawsuits. Additionally, legal battles over fair use can be long, costly, and risky. Another practical issue is that fair use rights can be limited by contracts (Carroll, 2019). Many websites, subscription services, and digital platforms include terms of use that ban TDM or ask users to give up their fair use rights. As a result, even content that's legally accessed may not be usable for TDM purposes.

In summary, while the fair use doctrine in the U.S. offers valuable flexibility for adapting to new technologies like TDM, its uncertain and case-by-case nature, high litigation costs, and vulnerability to contractual restrictions make it a complicated and sometimes unreliable option for those working in TDM. This stands in contrast to legal systems that offer clearer, specific rules and exceptions for such practices.

2. The United Kingdom: In the United Kingdom, TDM is specifically addressed in Section 29A of the Copyright, Designs and Patents Act 1988 (CDPA). This legal exception allows copyrighted works to be copied under strict conditions: the user must have lawful access to the material; the copying must be done solely for computational analysis of the work's content; and the purpose must be non-commercial research (Intellectual Property Office, 2014). If the research results are published and they include any part of the original work, the source must be properly

acknowledged, unless doing so is clearly impractical. One of the most important features of this law is that contractual terms cannot override it. This means content providers cannot block lawful TDM through license agreements or terms of service. While there have been efforts to expand TDM rights to support artificial intelligence, such as a 2022 proposal to allow TDM for any purpose, and a 2024 consultation on a "rights reservation" system, these proposals have not yet changed the current legal framework, which remains limited to non-commercial research (Scott, 2025). This can be seen that the UK's approach provides greater legal certainty for non-commercial research than Thai law does, and importantly, its statutory protection cannot be nullified by contracts which is a key advantage for academic researchers.

The UK's approach to TDM offers several strengths. Most notably, it provides clear legal certainty for researchers working within the scope of the exception. Unlike the U.S. fair use system, which depends on uncertain case-by-case judgments, UK researchers can conduct non-commercial TDM with more confidence that their actions are legal. Another major advantage is found in Section 29A (5), which prevents content providers from using contracts to restrict TDM activities that are otherwise allowed by law. This ensures that the exception remains useful in practice for those who have lawful access to the materials. In addition, the requirement for lawful access helps prevent the misuse of illegally obtained content.

Despite these benefits, the UK model also has some key limitations. The most significant issue is that the exception is limited to non-commercial research (Headdon, 2023). This restriction excludes most TDM activities conducted by

companies, especially those involved in developing commercial products, services, and AI technologies. At a time when AI and data-driven innovation are crucial for economic development, this narrow scope is often seen as a barrier to progress. As a result, critics argue that the law does not go far enough to support commercial innovation.

In summary, while the UK's TDM exception offers solid protection and clarity for non-commercial research, it falls short in supporting the broader, commercially-driven uses of TDM that are increasingly important in today's digital economy.

3. Japan: Japan has taken one of the most forward-thinking and inclusive approaches to TDM among the jurisdictions considered, reflecting a national policy aimed at boosting digital innovation and artificial intelligence development (OECD, 2015). The Japanese Copyright Act was first updated in 2009 to account for TDM, with major revisions made in 2018. Article 30-4 now serves as the key provision, allowing the use of copyrighted materials by any method, as long as the purpose is not to "enjoy or let others enjoy" the ideas or emotions expressed in the work. This "non-enjoyment" principle is central to the Japanese model and enables both commercial and non-commercial TDM without requiring prior permission (Ueno, 2021). Importantly, this exception cannot be overridden by contract, meaning copyright holders cannot restrict qualifying TDM uses through license agreements or terms of service (Fernández-Molina & De La Rosa, 2024).

Although Article 30-4 does not explicitly require "lawful access," this issue is indirectly addressed through a general condition that the use must not "unreasonably prejudice the interests of the copyright owner." Recent guidance, such as the 2024 report from Japan's Agency for Cultural Affairs (ACA), has clarified that TDM outputs should not unfairly harm or directly compete with the original work in ways that negatively impact the copyright holder's market (Grasser, 2024). It is clearly seen that Japan's permissive model, which supports both commercial and non-commercial TDM, stands in stark contrast to Thailand's framework, where even purely non-commercial TDM faces significant legal hurdles.

The Japanese approach provides several important benefits. Most notably, it offers strong legal clarity and support for innovation, allowing both commercial and academic users to engage in TDM with minimal legal risk. By explicitly allowing TDM without needing consent and by preventing copyright holders from opting out, Japan creates a favorable environment for investment in areas like AI, data science, and Big Data. The model also rests on a well-defined philosophical foundation: the "non-enjoyment" principle argues that TDM does not involve the kind of emotional or intellectual experience that copyright law is designed to protect. Instead, TDM is an analytical activity that extracts knowledge and insights, rather than replicating or competing with the original expressive content. In practical terms, the rule that contracts cannot override this exception ensures that users are not blocked by restrictive terms of use, something that remains a major barrier in jurisdictions like the United States.

However, Japan's permissive stance is not without controversy. One major criticism is the lack of control and compensation for creators, especially when their works are used in large-scale commercial TDM projects. Because copyright holders cannot opt out, some argue this undermines their autonomy and ability to decide how their works are used. There is also debate over whether commercial TDM should

involve compensation, even if it meets the non-enjoyment standard. Critics argue that since such activities can generate profits, creators should be entitled to some form of payment. However, the Japanese government maintains that as long as the use does not interfere with the original market, no exclusive rights or payment rights are justified (Ueno, 2021). Finally, while the general fairness requirement may discourage the use of illegally obtained content, the absence of a clear "lawful access" clause, unlike in laws such as the UK's, could be seen as a legal gap or point of ambiguity.

In summary, Japan's TDM framework stands out for its pro-innovation stance, strong legal clarity, and wide applicability. Still, it continues to face important debates around creator rights, market impacts, and the boundaries of fair exploitation.

4. China: China's approach to TDM under its copyright law reflects a civil law tradition that follows a closed-list system of exceptions (He, 2021). The Copyright Law of the People's Republic of China (CLC) 2020, in Article 24, outlines specific circumstances where copyrighted works can be used without the copyright owner's permission or payment. However, this list does not currently include a specific exception for TDM. Existing provisions, such as the one allowing use for "private study, research, or self-entertainment" (Article 24(1)) or for "classroom teaching or scientific research" involving small-scale reproduction (Article 24(6)), are generally too limited in scope (He, 2021). They are not suitable for the large-scale, often collaborative or commercially relevant nature of most TDM projects. Moreover, these exceptions are subject to the three-step test, which restricts exceptions to particular cases and requires that they do not

interfere with the normal exploitation of the work or unfairly harm the copyright holder's interests (Hue, 2022).

In 2020, a new clause was added under Article 24(1)(xiii), allowing future exceptions to be created through additional laws or administrative regulations. However, this clause functions more as a formal mechanism for gradual legal development than as an immediate solution for modern uses like TDM (He, 2021). In response to the limitations of the current law, some Chinese courts have begun to cautiously explore the idea of "transformative use", influenced by U.S. case law. For instance, in Wang Shen v. Beijing Guxiang Information Technology Co., Ltd. (2013), this concept was considered. Yet, because this idea is not grounded in Chinese statute, its application remains limited and inconsistent.

From the standpoint of traditional copyright protection, China's closed-list system offers one major advantage: strong protection for creators. By strictly defining permissible uses, the law provides clear boundaries and reduces the risk of unauthorized exploitation of creative works. It places a strong emphasis on safeguarding both the economic and moral rights of authors, aiming to prevent misuse or unintended re-use of copyrighted materials.

However, this same structure presents serious drawbacks for TDM and broader technological innovation. The most pressing issue is the lack of a legal pathway for conducting TDM. Without a dedicated exception, researchers and companies face a high risk of legal challenges, which creates hesitation and stifles innovation, especially in fields like artificial intelligence, which are central to China's economic and technological goals. The closed nature of the exceptions makes the system slow to respond to new

technologies, as changes require formal legal or administrative amendments. It can be seen that China's restrictive, closed-list framework is structurally similar to Thailand's current legal landscape, where the absence of a dedicated TDM exception creates significant barriers to data-driven innovation. Additionally, the option to introduce new exceptions through Article 24(1)(xiii) is a structured and deliberate process, offering no quick relief for emerging needs like TDM.

In conclusion, the current exceptions in Article 24 are too limited for typical TDM activities, restricting content use, prohibiting commercial applications, and limiting result sharing. Judicial attempts to apply "transformative use" lack legal clarity, offering little reliable protection. This highlights a key tension between China's strict copyright laws and its AI ambitions. To support digital innovation, a more flexible and adaptive legal framework for TDM is needed

Taken together, these four jurisdictions illustrate a spectrum of legal approaches to TDM. They range from the flexible U.S. model and Japan's permissive commercial exception to the UK's narrowly scoped right and China's closed-list framework. Thailand's law currently aligns most closely with China's in its restrictiveness. To foster innovation, Thailand would benefit from studying the clear, reservation protection statutory clarity offered by the UK and Japan

Legal Reform for TDM in Thailand: Justification and Proposal

The current legal framework under the Thai Copyright Act B.E. 2537 (1994), particularly Section 32, is not well-suited to support modern TDM activities. The general exceptions currently available are limited by narrow court interpretations, especially the strict "not for profit" requirement and the view that extensive copying inherently infringes on a copyright holder's right to commercially exploit their work. These limitations create serious legal uncertainty and practical challenges for researchers and developers engaging in TDM, effectively restricting both non-commercial and commercial applications.

This legal gap is at odds with Thailand's broader national strategies, such as the "Digital Thailand" initiative and the National AI Strategy (2022–2027), which aim to develop a strong data-driven economy and position the country as a leader in artificial intelligence within the region (Ministry of Digital Economy and Society, 2016; The Nation, 2025). To achieve these goals, a legal framework that supports both academic and commercial innovation is essential. Including commercial purposes within a TDM exception is crucial for fostering a vibrant ecosystem of AI businesses, encouraging public-private R&D partnerships, and ensuring Thailand can compete regionally and globally. Restricting TDM to non-commercial activities would stifle the development of practical applications and limit the economic benefits that data-driven technologies can offer.

A foundational element of the proposed reform is the prerequisite of "lawful access." This principle ensures that the TDM exception can only be invoked when the user has legally obtained the copyrighted materials, such as through a library subscription, purchase, or an open-access license. By requiring lawful access, the framework prevents the exception from being used to legitimize pirated or illegally obtained content, aligning the reform with ethical standards and the legal approaches of the UK and the US.

Central to the proposal is the "non-enjoyment" principle, as outlined in clause (1)(a). This provision, inspired by Japan's innovative legal framework, permits TDM on the condition that the primary purpose is computational analysis to derive insights, not to experience or "enjoy" the expressive content of the work itself. This distinction is critical because it separates non-consumptive, transformative uses like TDM from activities that would typically compete with the copyright holder's primary market, providing a strong and clear justification for the exception.

Furthermore, the proposed exception is intentionally designed to align with Thailand's international obligations. The structure, particularly clause (1)(b), incorporates the principles of the internationally recognized "three-step test" found in the Berne Convention and the TRIPS Agreement. This ensures that the exception is balanced, applying only to special cases that do not conflict with the normal exploitation of the work or unreasonably prejudice the legitimate interests of the rights holder, thereby demonstrating compliance with global copyright standards.

A critical component of this reform is the prevention of contractual override. The proposed clause (3) ensures that the statutory exception cannot be nullified by restrictive licensing agreements, which are often non-negotiable. This clause aligns with international best practices in jurisdictions like the UK and Japan, which recognize that without such protection, a TDM exception would be ineffective in practice. It ensures that the right to perform TDM is meaningful and cannot be circumvented by private contracts that would otherwise hinder lawful research and innovation.

Finally, the proposal incorporates essential post-analysis safeguards in clauses (4) and (5) to ensure the exception is balanced and respects the rights of creators. Clause (4) addresses a primary concern of copyright holders by mandating that copies made for TDM must be stored securely and not be made available to the public in a way that could substitute for the original work. This crucial safeguard prevents the TDM corpus from becoming an infringing public library and ensures its use is strictly limited to computational analysis. Clause (5) further reinforces this balance by requiring appropriate acknowledgement if the published results contain substantial parts of the original expressive works which represent the core information of the works. This upholds the moral rights of authors and aligns with responsible research practices, mirroring requirements in other legal frameworks like the UK's.

For the aforementioned rationales, the author proposes that Thailand adopt a specific legal exception for TDM. The following draft outlines a suggested new section of the law:

"Section XX: Exploitation of Works for Computational Analysis

(1) The reproduction of a work to which a person has lawful access, made for the purpose of computational analysis of data, text, sounds, images, or other elements contained within such work including the use of automated techniques to process and analyze content, such as for Text and Data Mining, AI model training, machine learning, and statistical analysis, shall not be deemed an infringement of copyright, provided that:

(a) the primary purpose of such reproduction or adaptation is not to enjoy or cause another person to enjoy the expressive

elements of the work itself, but rather to derive information, patterns, trends, or other insights through computational means; and

(b) such exploitation does not conflict with a normal exploitation of the copyrighted work by the owner of copyright and does not unreasonably prejudice the legitimate interests of the owner of copyright, taking into account the nature of the copyrighted work, the purpose and character of the exploitation, and the extent of the reproduction in relation to the computational analysis undertaken.

(2) The provisions of this section shall apply to exploitation for both non-commercial and commercial purposes.

(3) By virtue of this section, any contractual term that purports to prohibit or restrict the making of a reproduction of a work which would not infringe copyright, shall be unenforceable.

(4) Copies of works made under this section must be stored securely and must not be communicated to the public or otherwise made available in a manner that would substitute for the original work or serve as a means for its direct enjoyment by the public, except as may be permitted by authorization from the copyright owner.

(5) Where the results of computational analysis conducted under this section are published or otherwise communicated to the public, and such results include direct quotations or expressive reproductions from the original works that constitute a substantial part of the original work, appropriate acknowledgement of the source of such works shall be made, unless this would be impossible for reasons of practicality or otherwise."

Conclusion

TDM is essential for digital innovation, but copyright laws often hinder its use. This article shows that while countries like the US, UK, Japan, and China address TDM in different ways, Thailand's current Copyright Act B.E. 2537 (1994) and its strict interpretation create major legal barriers. These restrictions limit both commercial and non-commercial TDM, slowing Thailand's progress toward its digital and AI goals. Given this, the study concludes that Thailand urgently needs a clear, specific TDM exception. Such reform would reduce legal uncertainty, support innovation, and benefit a wide range of stakeholders, including academic researchers, AI developers, small enterprises, and public institutions, by enabling lawful, responsible access to data. Importantly, a well-designed exception would preserve the legitimate interests of copyright holders while fostering a data-driven economy. By aligning Thai law with global trends and international commitments like the TRIPS Agreement and the Berne Convention, this reform would help balance copyright protection with the country's development goals. In the long term, a dedicated TDM exception can serve as a catalyst for responsible AI innovation, boost research productivity, and position Thailand as a regional hub for ethical and data-driven technologies.

References

Carroll, M. W. (2019). *Copyright and the progress of science: Why text and data mining Is lawful (SSRN Scholarly Paper No. 3531231). Social Science Research Network.* Retrieved from https://papers.ssrn.com/abstract=3531231

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine, 17*(3), 37–54. Doi: https://doi.org/10.1609/aimag.v17i3.1230

Fernández-Molina, J.-C., & De La Rosa, F. E. (2024). Copyright and text and data mining: Is the current legislation sufficient and adequate?. *portal: Libraries and the academy, 24*(3), 653–672. Doi: https://doi.org/10.1353/pla.2024.a931775

Grasser, S. W., & Joseph. (2024). *Japan's new draft guidelines on AI and copyright: Is It really ok to train ai using pirated materials? Privacy World.* Retrieved from https://www.privacyworld.blog/2024/03/japans-new-draft-guidelines-on-ai-and-copyright-is-it-really-ok-to-train-ai-using-pirated-materials/

Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text mining in big data analytics. *Big data and cognitive computing, 4*(1), 1. Doi: https://doi.org/10.3390/bdcc4010001

Headdon, T. (2023). *The text and data mining copyright exception in the UK "for the sole purpose of research for a non-commercial purpose" What does it cover? Bristows*. Retrieved from https://www.bristows.com/news/the-text-and-data-mining-copyright-exception-in-the-uk-for-the-sole-purpose-of-research-for-a-non-commercial-purpose-what-does-it-cover/

He, T. (2021). *Copyright exceptions reform and ai data analysis in China a modest Proposal. In J.-A. Lee, R. Hilty, & K.-C. Liu (Eds.), Artificial intelligence and intellectual property pp. 196–220).* Oxford University PressOxford. Doi: https://doi.org/10.1093/oso/9780198870944.003.0010

He, T. (2022). *The copyright liimitations of the 2020 copyright Law of China: A Satisfactory compromise? the copyright society*. Retrieved from https://copyrightsociety.org/journal-entries/the-copyright-limitations-of-the-2020-copyright-law-of-china-a-satisfactory-compromise/

Hue, J. (2022). *Copyright exceptions for text and data mining in China: Inspiration from transformative use the copyright society.* Retrieved from https://copyrightsociety.org/journal-entries/copyright-exceptions-for-text-and-data-mining-in-china-inspiration-from-transformative-use/

Intellectual Property Office. (2014). *Exceptions to copyright-guidance for creators and copyright owners.* n.p.

Losiewicz, P., Oard, D. W., & Kostoff, R. N. (2000). Textual data mining to support science and technology management. *Journal of intelligent information systems, 15*(2), 99–119. Doi: https://doi.org/10.1023/A:1008777222412

Ministry of Digital Economy & Society. (2016). *Thailand digital economy and society development plan.* Retrieved from https://dig.watch/resource/thailand-digital-economy-and-society-development-plan

Natasha, K. C. K. (2022). *The role of copyright law in text and data mining research.* Retrieved from https://cipit.strathmore.edu/the-role-of-copyright-law-in-text-and-data-mining-research/

The Nation. (2025). *Thailand outlines ambitious AI strategy to become regional hub*. Retrieved from https://www.nationthailand.com/business/tech/40049494

OECD. (2013). *Exploring data-driven innovation as a new source of growth: Mapping the policy issues raised by "big data" (oecd digital economy papers no. 222; oecd digital economy papers, vol. 222)*. Doi: https://doi.org/10.1787/5k47zw3fcp43-en

OECD. (2015). *Data-driven innovation: Big data for growth and well-being.* OECD. Doi: https://doi.org/10.1787/9789264229358-en

Rosati, E. (2019). *Copyright as an obstacle or an enabler? a European perspective on text and data mining and its role in the development of AI creativity (ssrn scholarly paper no. 3452376). social science research network.* Retrieved from https://papers.ssrn.com/abstract=3452376

Samuelson, P. (2009). Unbundling fair uses. *Fordham law review, 77*(5), 25-37.

Samuelson, P. (2015). *Justifications for copyright limitations & exceptions (ssrn scholarly paper no. 2476669). social science research network.* Retrieved from https://papers.ssrn.com/abstract=2476669

Scott, E. (2025). *Copyright and artificial intelligence: Impact on creative industries.* Retrieved from https://lordslibrary.parliament.uk/copyright-and-artificial-intelligence-impact-on-creative-industries/

Ueno, T. (2021). The flexible copyright exception for 'non-enjoyment' purposes-recent amendment in Japan and its implication. *GRUR International, 70*(2), 145–152. Doi: https://doi.org/10.1093/grurint/ikaa184

Weber, N., & Locke, B. (2022). *Ethics of open data.* arXiv. Doi: https://doi.org/10.48550/ARXIV.2205.10402