

AI Training and Copyright Law: An Analysis of the Thomson Reuters v. Ross Intelligence

by Mohammad Ataul Karim, LL.M¹

I. Introduction

In *Thomson Reuters v. Ross Intelligence*,² Judge Stephanos Bibas of the United States District Court for the District of Delaware recently issued a summary judgment regarding AI training, copyright infringement, and the fair use doctrine. It is now pending before the United States Court of Appeals for the Third Circuit. According to Judge Stephanos Bibas, Ross Intelligence infringed on the copyright of the Westlaw legal database by copying the same for the training of its AI search tool. The decision is considered the first of its kind concerning AI training, copyright and fair use. Currently, several copyright litigations over training generative AI are pending. None of them has been decided yet. Therefore, this decision warrants further analysis, as it may offer valuable insights for AI training, copyright infringement, and copyright exceptions. This paper aims to do so by analysing this decision. The paper is designed in the following way. Parts II and III of this paper outline the context of the dispute and the court's decision. Part IV presents an analysis of the decision, and Part V highlights some key insights from it. Part VI concludes the discussion.

II. A Factual Overview of the Dispute

Thomson Reuters owns Westlaw, a renowned legal resource database that compiles U.S. state and federal statutes, regulations, case law, and journals, among other materials.³ In particular, two elements of the Westlaw database, the headnotes and the Key Number System (KNS), are central to this dispute. First, a headnote is a concise summary of a key point of law extracted from a lengthy judicial opinion.⁴ In other words, a headnote precisely explains the lengthy judicial opinion, key legal points, and holdings. It helps readers quickly grasp a specific legal issue without going into details. Second, KNS is a numerical taxonomy for organising contents.

¹ Mohammad Ataul Karim, LL.M is DPhil in Law Candidate (Intellectual Property) at the Faculty of Law, University of Oxford. The views expressed in this article are those of the author and do not necessarily represent those of his former employers or Oxford university. All links have been accessed on 2 July 2025.

² THOMSON REUTERS ENTERPRISE CENTRE GMBH and WEST PUBLISHING CORP., Plaintiffs, v. ROSS INTELLIGENCE INC., Defendant. IN THE UNITED STATES DISTRICT COURT FOR THE DISTRICT OF DELAWARE (No. 1:20-cv-613-SB) (February 11, 2025), https://www.ded.uscourts.gov/sites/ded/files/opinions/20-613_5.pdf

³ *ibid*, 2

⁴ *ibid*, 7

It categorises the materials in the database into different buckets for the convenience of the readers. Thus, it is a numerically convenient way to organise the contents in the database.

Ross Intelligence, a competitor of Westlaw database, wanted to design and organise a similar legal research platform using an artificial intelligence (AI) search tool.⁵ It wanted to use Westlaw's database (as input) to train its AI search tool. Thomson Reuters refused to license Ross Intelligence for training its AI search tool.⁶ Ross looked for an alternative source to train its AI search tool. Consequently, it agreed with LegalEase, a third-party entity, to use so-called Bulk Memos. It brought 25,000 Bulk Memos. Bulk Memos comprise a compilation of lawyers' questions with good and bad answers.⁷ LegalEase provides a guide explaining how to create those questions via Westlaw's headnotes. It also clarifies that lawyers should not just copy and paste headnotes into the questions.⁸ Ross Intelligence trained its AI search tool using Bulk Memos, which were originally developed based on Westlaw's headnotes.⁹ Therefore, Thomson Reuters sued Ross Intelligence for training its AI search tool by infringing on copyrighted materials.

III. Decision of the Court

The key issue in the court was whether the 'Bulk Memo Questions' copied Thomson Reuters's copyrightable headnotes or uncopyrightable judicial opinions. The court has adopted an innovative way. It constructed the following sample based on *the Feist Case* situation to run a side-by-side analysis of the disputed Issues:¹⁰ Feist situation was adopted draw an analogy with this case by outlining the nature of Bulk memo question, Westlaw headnote, and case opinion.

Bulk Memo Question	Westlaw Headnote	Case Opinion
Does originality for copyright purposes mean that the work was independently created and has some minimal degree of creativity?	Originality, for copyright purposes, means that the work was independently created and has some minimal degree of creativity.	Original, as the term is used in copyright, means only that the work was independently created by the author (as opposed to copied from other works), and that it

⁵ *ibid*, 3

⁶ *ibid*

⁷ *ibid*

⁸ *ibid*

⁹ *ibid*

¹⁰ *Feist Publications, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 345 (1991), *ibid*, 4

		possesses at least some minimal degree of creativity.
--	--	---

The case opinions are generally unprotected whereas headnotes can be protected by copyright. The court held that: a) Ross Intelligence is liable for copying 2,243 headnotes for its AI training and b) All defences of Ross, namely innocent infringement, copyright misuse, merger, and fair use, failed.¹¹

IV. An Analysis of the Decision

A. Validity of Copyright

The first issue is the validity of the copyright. The validity of Westlaw's copyright for its content was presumed due to its registration.¹² Copyright protection is provided for original expressions. The *Feist* rule was considered by the court regarding the originality threshold which is “extremely low, requiring only some minimal degree of creativity...some creative spark”.¹³ A case opinion (see the table above) is not generally copyrightable. However, editorial notes and comments on a case opinion may be copyrighted. For this case, a headnote can be creative and copyrightable because it involves “distilling, synthesising, or explaining part of a judicial opinion.”¹⁴ A compilation of headnotes can be original “if the compiler makes choices regarding selection and arrangement” by using “a minimum degree of creativity.”¹⁵ Thus, the court held that headnotes, individually and as a compilation, clear the bar of the originality requirement for copyright.

The originality of the KNS was challenged because they divided the content based on the common legal doctrinal questions.¹⁶

Arguably, they are not novel and thus lack originality. The court rejected this and opined that KNS possesses a ‘minimum spark of originality.’¹⁷ Thomson Reuters's labour, investment, and hard work in creating KNS and its utility for organising the legal database are immaterial for the purpose of originality. Instead, the key consideration is whether KNS was created independently. There are many ways to organise legal topics in a database. Among the many options, Thomson Reuters adopted KNS, which is sufficient to pass the condition for

¹¹ Ibid, 5

¹² The registration is prima facie evidence of validity if it is done before or within five years after the first publication of work. Section 410 (c), 17 USC

¹³ *Feist Publications, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 345 (1991)

¹⁴ *ibid*, 7

¹⁵ *ibid*

¹⁶ *ibid*, 8

¹⁷ *ibid*

independent creation.¹⁸ In summary, the court concludes that both headnotes and KNS are original, copyrightable, and the copyright over Westlaw's contents was presumed to be valid due to the registration.

B. Infringement of Copyrights

The infringement analysis examines whether there is a) actual copying and b) substantial similarity between the protected works and the alleged infringing materials.¹⁹

First, actual copying occurs when “the defendant did, in fact, use the copyrighted work in creating his own.”²⁰ It can be proved by direct evidence—the defendant copied the work—or indirect evidence—the defendant had access to the work and produced something similar to it (probative similarity).²¹

The court analysed the 2,830 Bulk Memo questions, headnotes, and judicial opinions. It was not difficult for the court to find actual copying. First, Ross Intelligence did not deny that LegalEase had access to Westlaw and used it to create BulkMemos.²² Moreover, Ross’s counsel submitted that some bulk memos were discarded during the AI training process. However, it was confirmed twice that 80% of them were used for initial AI training, and the remaining 20% were used for validation.²³ The court found the actual copying of 2,243 headnotes out of 2,830 Westlaw’s headnotes.²⁴

Second, the court addressed the issue of substantial similarity. For the later work to be substantially similar, it must materially appropriate the earlier work.²⁵ The court considered whether an ordinary user of a legal database would find it substantially similar.²⁶ Judge Bibas, as a lawyer and judge, considered himself an ordinary user of the Westlaw database. He concluded that Bulk Memo questions are substantially similar to 2,243 Westlaw’s headnotes.²⁷

¹⁸ *ibid*

¹⁹ *Ibid*, 11

²⁰ *Tanksley v. Daniels*, 902 F.3d 165, 173 (3d Cir. 2018)

²¹ *THOMSON REUTERS ENTERPRISE CENTRE GMBH and WEST PUBLISHING CORP., Plaintiffs, v. ROSS INTELLIGENCE INC., Defendant*, 11

²² *ibid*, 12

²³ *ibid*

²⁴ *ibid*

²⁵ *Tanksley*, 902 F.3d at 173. *Ibid*, 13

²⁶ *Dam Things*, 290 F.3d at 562. *ibid*, 13

²⁷ *ibid*, 14

In summary, the court found a) the actual copying of 2,243 Westlaw's copyrighted headnotes, and b) substantial similarity between the said headnotes and Bulk Memo questions. Thus, the court awarded a summary judgment against Ross Intelligence for copyright infringement.

C. Defences of Ross Intelligence

After an infringement is determined by the court the defendant could bring relevant defences to defeat the infringement. The question was whether Ross Intelligence had any valid defence(s). It was held that Ross's defences, namely innocent infringement, copyright misuse, merger, and fair use, failed. Fair use was the key defence; therefore, it is considered in detail in comparison to other defences.

1. Innocent Infringement, Copyright Misuse, Merger Defence

First, Ross unsuccessfully argued for innocent infringement. Innocent infringement happens when someone infringes copyright without actual and constructive knowledge. Since Westlaw's headnotes bear a copyright notice, innocent infringement cannot be claimed.²⁸ Second, the court rejected the doctrine of copyright misuse. Ross claims that Thomson Reuters misused copyright by refusing to license, which is anti-competitive. The court opined that Ross failed to prove the misuse of copyright for stifling competition.²⁹ Third, Ross argued that merging ideas (judicial opinion) and expressions (Westlaw's headnotes and KNS) makes them uncopyrightable (merger defence).³⁰ The court rejected the merger defence by deducing that there were many ways to express legal points from judicial opinions, and Westlaw successfully did that through headnotes and KNS.³¹

2. Analysis of the Fair Use Doctrine

Section 107 of the U.S. Copyright Act of 1976 requires a four-factor analysis for the "fair use" defence.³² *The first factor* is the purpose and character of the use, including whether it is for

²⁸ *ibid*,

²⁹ *ibid*,15

³⁰ *Educ. Testing Servs. v. Katzman*, 793 F.2d 533, 539 (3d Cir. 1986), *ibid*, 15

³¹ *THOMSON REUTERS ENTERPRISE CENTRE GMBH and WEST PUBLISH ING CORP., Plaintiffs, v. ROSS INTELLIGENCE INC., Defendant*. P. 15

³² "Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or

commercial or nonprofit educational purposes.³³ Ross's use was commercial and not transformative.³⁴ Ross's use of the 'headnotes' for commercial nature was undisputed. The headnotes were used to train an AI tool for developing a commercial legal database, similar to Thomson Reuters' commercial use. Moreover, the headnotes were copied to train an AI search tool, which was used to create a legal research database to compete with Thomson Reuters. Thus, such use is not transformative because it lacks a further purpose or distinct character from that of Thomson Reuters.³⁵

The second factor is the nature of the copyrighted work.³⁶ Regarding this factor, the court considers the degree of creativity inherent in the work. In its opinion, more creative works receive greater protection.³⁷ As a database, Westlaw is not particularly creative. It, however, meets the minimum threshold of copyrightability. Although headnotes require some editorial contributions, they are not at the same level as those required for novels or artistic works. Moreover, the KNS is a factual compilation; hence, space for creativity is limited. In summary, the second factor favours Ross Intelligence but does not significantly impact fair use determination.³⁸

The third factor is "the amount and substantiality of the portion used in relation to the copyrighted work as a whole".³⁹ The court inquired whether copying was reasonable, considering the quantity and quality of the copied materials.⁴⁰ The key is that "the copier must not take the heart of the work".⁴¹ Ross successfully argued that output from the AI training

research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include—

(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;

(2) the nature of the copyrighted work;

(3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and

(4) the effect of the use upon the potential market for or value of the copyrighted work."

³³ 17 U.S.C. § 107(1)

³⁴ *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 529–31 (2023), THOMSON REUTERS ENTERPRISE CENTRE GMBH and WEST PUBLISHING CORP., Plaintiffs, v. ROSS INTELLIGENCE INC., Defendant. P. 16

³⁵ THOMSON REUTERS ENTERPRISE CENTRE GMBH and WEST PUBLISHING CORP., Plaintiffs, v. ROSS INTELLIGENCE INC., Defendant. P. 17

³⁶ 17 U.S.C. § 107(2)

³⁷ THOMSON REUTERS ENTERPRISE CENTRE GMBH and WEST PUBLISHING CORP., Plaintiffs, v. ROSS INTELLIGENCE INC., Defendant. P. 20

³⁸ 17 U.S.C. § 401(d), *ibid.*

³⁹ 17 U.S.C. § 107(3)

⁴⁰ *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 587 (1994) THOMSON REUTERS ENTERPRISE CENTRE GMBH and WEST PUBLISHING CORP., Plaintiffs, v. ROSS INTELLIGENCE INC., Defendant. P. 21

⁴¹ THOMSON REUTERS ENTERPRISE CENTRE GMBH and WEST PUBLISHING CORP., Plaintiffs, v. ROSS INTELLIGENCE INC., Defendant. P. 21

would only communicate judicial opinions (unprotected parts) to the end users, not Westlaw's headnotes (protected parts).⁴² The court accepted this. Ross unsuccessfully argued that it copied only a small percentage of Westlaw's total headnotes for AI training, compared to the entire database; thus, it should be permitted.⁴³ In summary, the court concluded that the third factor favours Ross.

The fourth factor is “the effect of the use upon the potential market for or value of the copyrighted work”.⁴⁴ This factor was considered as the single most important factor for determining fair use.⁴⁵ The court identified two markets: a) the original market for legal research platforms and b) the potential derivative data market for training legal AI.⁴⁶ In assessing this factor, the court examined the three issues.

First, Ross's use would be transformative if it creates a brand-new research platform serving a purpose different from that of Westlaw.⁴⁷ It means Ross's database would not be a market substitute for Westlaw. It was already evident in the analysis of the first factor that Ross's use is not transformative, as it serves the same purpose as Westlaw.

Second, how Ross's copying potentially affects the market for data used in AI training. In assessing this, it was relevant whether Thomson Reuters would use, sell, or license its data for AI training. The court concluded that the potential negative effect on the derivative market—the data market for AI training—is sufficient, regardless of whether Thomson Reuters has used the contested data to train its own search tool.⁴⁸

Finally, the question was whether there is any public interest that can be better served by protecting the creator or the copier.⁴⁹ Indeed, the court acknowledges a competing public interest for access to the law. However, this does not restrict Thomson Reuters' commercial interests, which are protected by copyright. Moreover, legal opinions are freely available. Thus, Thomson Reuters wins in factor four.

⁴² *ibid*, 21

⁴³ THOMSON REUTERS ENTERPRISE CENTRE GMBH and WEST PUBLISHING CORP., Plaintiffs, v. ROSS INTELLIGENCE INC., Defendant. P. 21

⁴⁴ 17 U.S.C. § 107(4)

⁴⁵ THOMSON REUTERS ENTERPRISE CENTRE GMBH and WEST PUBLISHING CORP., Plaintiffs, v. ROSS INTELLIGENCE INC., Defendant. P. 21

⁴⁶ *ibid*, 22

⁴⁷ *ibid*

⁴⁸ *ibid*

⁴⁹ *ibid*

In summary, the court held that factors one and four support Thomson Reuters, while factors two and three favour Ross Intelligence.⁵⁰ In the balancing analysis, “factor two matters less than the others, and factor four matters more”.⁵¹ Thus, Ross's fair use defence fails.

V. Some Insights from this Decision

First, this is a district-level decision, which should not be treated as conclusive for AI training and fair use. Second, this case is not about training generative AI but rather a specific AI search tool for databases. Therefore, this decision is unlikely to set a precedent for the pending cases revolving around generative AI training and copyright law. Again, there are different types of AI systems and different stages of AI training. Thus, specific aspects of copyright would be relevant to the *pre-training*, *training*, and *post-training activities* of a generative AI model or system.⁵² Moreover, the relevance of copyright and fair use for specific AI training would be a case-by-case analysis.

In addition, the tests for AI training and copyright exceptions in the EU and UK would differ to the fair use approach in the US. The EU and UK have narrow text and data mining (TDM) exceptions.⁵³ There is no conclusive judicial determination of AI training and TDM within these jurisdictions. So far, the LAION Case in Germany has only addressed a specific issue of AI training and TDM, namely, listing and sharing publicly available data sources for AI training.⁵⁴ Thus, neither fair use nor TDM exceptions should be seen as synonymous with AI training, at least to date. Still, these exceptions provide the basis of many defence arguments in pending judgements against AI developers.

Moreover, training data for AI is normally substitutable. In other words, there are alternative data sources for AI training. In cases like this, refusal to license specific data is unlikely to constitute an abuse of copyright unless there is a specific competition or antitrust issue.⁵⁵

⁵⁰ *ibid*, 23

⁵¹ *ibid*

⁵² Copyright and Generative AI: Opinion of the European Copyright Society, January 2025, 6 https://europeancopyrightsociety.org/wp-content/uploads/2025/02/ecs_opinion_genai_january2025.pdf

⁵³ The Copyright, Designs and Patents Act (the CDPA) 1988, s. 29A, The Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, OJ L 130, 17.5.2019, p. 92–125, ELI: [http://data.europa.eu/eli/dir/2019/790/oj\(1\)](http://data.europa.eu/eli/dir/2019/790/oj(1)), (The DSM Directive), Art. 3 and 4.

⁵⁴ Germany - Hamburg District Court (Landgericht Hamburg), As. 310 O.227/23, LAION v Robert Kneschke, [27 September 24] <<https://www.euipo.europa.eu/en/law/recent-case-law/germany-hamburg-district-court-310-o-22723-laion-v-robert-kneschke>>

⁵⁵ See for similar reasoning, Joined Case C-241/91 P and C-242/91 P, TRE and ITP v. Commission, (1995) ECR I-743; Case C-418/01, IMS Health, (2004) ECR I-5039

Lastly, obtaining copyright clearance from third parties is important for AI training. It is crucial for the legal certainty of AI training activities and ensuring fair and equitable remuneration for rights holders. Thus, data brokers, providers, and users must clear the existing rights (including copyright). They cannot invoke the innocent infringement defence by relying on the data brokers or data providers. For example, in this case, the data provider (LegalEase) failed to clear the existing right (copyright), and Ross Intelligence did not take sufficient steps to invoke an innocent infringement defence.

VI. Conclusion

This decision may not establish an authoritative precedent, particularly in the context of generative AI training and copyright law. Moreover, copyright infringement and fair use tests are subject to a case-by-case analysis. Therefore, it is challenging to draw any general conclusion. Furthermore, the legal tests in the EU and the UK are different. Unlike fair use, they have narrow exceptions, such as TDM exceptions. However, this decision provides some insights. There is no blanket exemption for using copyrighted materials without prior authorisation of the rights holder to train and develop AI systems or tools. The public interest in access to law or legal information does not always supersede the commercial interest of the copyright owner. A broad claim of copyright misuse due to refusal to license, without demonstrating relevant competition or antitrust effects, is not sustainable. The rights clearance is one of the key issues to avoid litigation and ensure legal certainty and fair remuneration of the rights owners. It is to be seen whether United States Court of Appeals for the Third Circuit uphold or reject the decision of the District Court.