Proceedings of the 4th International Conference on Global Politics and Socio-Humanities
DOI: 10.54254/2753-7048/2026.35831

Copyright Rules for Commercial Chess Game Databases: A Balance of AI Commercialization and Chess Players' Property Rights

Yifei Wang

Department of Law, Northwest Normal University, Lanzhou, China
Faye20051117@outlook.com

Abstract. Driven by the rapid development of generative artificial intelligence, commercial chess AI platforms widely crawl grandmasters'tournament game notations without formal authorization, triggering cross-border copyright disputes. Taking unlicensed chess training data conflicts as the research object, this paper adopts comparative law, case analysis and normative interpretation methods to compare copyright rules in China and the United States from four dimensions: originality of competitive game records, the scope of fair use, compensation standards for mass infringement, and centralized licensing mechanisms. Both jurisdictions recognize that serialized move sequences created by grandmasters through independent strategic judgment possess copyright originality. Nevertheless, the two legal systems share four structural defects: commercial text and data mining cannot qualify for statutory fair use; single-work compensation standards fail to address large-scale database infringement; individual one-to-one licensing brings excessive cross-border transaction costs; there are no targeted norms for sports data applied to AI training. To reconcile industrial innovation of chess AI and copyright interests of athletes, this paper proposes a dual-track cross-border governance framework: tiered differentiated compensation rules under national copyright legislation, and a unified global collective licensing platform managed by FIDE. This institutional design realizes equitable benefit distribution between AI developers and chess grandmasters, and can be extended to resolve similar data copyright disputes in all competitive sports industries.

Keywords: chess AI, chess game records, copyright originality, collective licensing, cross-border copyright governance

1. Introduction

The intersection of chess algorithm technology and intellectual property rights can be traced back to the landmark historical event that IBM Deep Blue defeated world champion Garry Kasparov in 1997. At that time, Deep Blue's core training corpus contained more than 700,000 historical grandmaster game records, all of which were captured through unrestricted collection without any copyright authorization or royalty payment to the originating players. This incident not only shocked the global chess community but also exposed the long-hidden institutional defects in the field of

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2026.35831

chess data utilization: the intellectual achievements created by players became free raw materials for technological enterprises to realize commercial profits without any corresponding interest feedback mechanism.

Over the past two decades, with the popularization of crawler technology and the breakthrough of deep learning algorithms, the industrial chain of commercial chess AI has matured rapidly. Well-known overseas platforms such as ChessBase and Chess.com, together with numerous domestic Chinese chess engine service providers, deploy automated crawler scripts to continuously grab full-game notations from global live broadcast rooms, fan forums and official tournament archives 24 hours a day. These serialized move records carry grandmasters'decades of accumulated competitive experience, personalized opening and endgame systems, as well as dynamic risk judgment logic formed under high-stakes match pressure, and have become irreplaceable core training raw materials for subscription analysis software, paid one-on-one tutoring courses and professional competition auxiliary AI. Paradoxically, grandmasters who devote a great deal of time, energy and intellectual labor to create such core data assets cannot share any commercial dividends generated by AI products relying entirely on their game records.

China and the United States represent the world's most influential continental copyright system and common law copyright system respectively, with relatively complete legislative and judicial systems, yet neither country has issued targeted balanced regulatory provisions to resolve the conflicts between chess AI commercialization and grandmasters' copyright property rights. The legal disputes triggered by mass scraping of chess data involve a series of core basic copyright theories, including the identification standard of originality of competitive sports outputs, the boundary of fair use applicable to commercial TDM, proportional damage calculation standards for scattered mass infringement, and operable institutional paths for large-scale centralized licensing. At present, the existing academic research has obvious fragmentation defects: foreign intellectual property scholars mostly only discuss whether single chess moves can be protected by copyright, ignoring the whole serialized game records with complete creative logic; domestic sports law and copyright literature mainly focus on the copyright protection of sports live video and audio works, while there are few comparative legal studies specially targeting textual chess game training data. Based on the obvious research gap in the existing academic circle, this paper sorts out the parallel copyright rules of China and the United States, summarizes four major institutional loopholes that jointly form industry-wide legal gray zones, and designs a complete dual-track cross-border governance scheme combining domestic legislative optimization and international sports organization self-regulation. The research conclusions can be borrowed and adjusted to solve similar AI training data copyright disputes in basketball, football, table tennis, e-sports and other competitive sports industries worldwide, providing universal normative reference for balancing athlete rights and technological industrial development..

2. Legal basis of copyright protection for chess game records

2.1. Basic concepts and core legal connotations

A chess game database refers to a standardized, machine-readable digital archive that completely records all move notations of formal over-the-board and online tournaments officially recognized by FIDE. All game data are stored in unified text formats to facilitate batch parsing, cloud storage and iterative training of machine learning models. Chess game notation is the tangible fixation carrier of grandmasters' instantaneous strategic choices in competitive matches. Different from fixed printed opening books that only summarize theoretical routines, every complete two-player tournament

DOI: 10.54254/2753-7048/2026.35831

game forms an exclusive interactive tactical chain through continuous mutual restraint and real-time risk weighing of both sides. Each match cannot be simply copied by pre-set algorithms, and it is the concrete externalization of long-term independent intellectual labor and unique competitive styles of professional players.

Chess players' copyright property rights refer to the bundle of exclusive commercial exploitation rights conferred by copyright law on their independently created match move sequences. Specifically, the scope of rights includes the right to authorize third-party AI companies to crawl, copy, archive and conduct algorithm training through TDM technology, as well as the statutory right to obtain proportional royalty remuneration from all profit-making commercial utilization acts. A clear theoretical dividing line must be drawn between idea and expression: the unified basic chess rules including board layout, piece movement tracks and winning and losing judgment conditions belong to universal public functional ideas, which cannot be monopolized by any individual through copyright protection; on the contrary, the individualized, match-exclusive continuous move chains formed by real-time confrontation and independent judgment of grandmasters contain unique human creative selection, and thus belong to original expressive works eligible for full copyright protection.

In addition, it is necessary to distinguish chess game records from pure sports statistical data such as winning rates, average scores and game durations. The latter only reflect objective numerical results without human subjective creative judgment, while the game notation records the whole dynamic decision-making process of both players, with obvious personality and creative attributes, which is the core premise for it to obtain copyright relief when infringed.

2.2. Jurisprudential foundation of chess game copyright protection under U.S. law

The U.S. copyright protection logic for chess game records is rooted in the classic idea-expression dichotomy stipulated in 17 U.S.C.§102(b), which clearly separates unprotectable abstract functional ideas and fixed original expressive works eligible for copyright. Combined with the landmark case Feist Publications, Inc. v. Rural Tel. Serv. Co., the U.S. judicial circle has formed a unified originality judgment standard: works only need to reflect minimal human creative selection, without the requirement of artistic aesthetic value. General chess operation rules are public ideas accessible to all chess practitioners, so they are excluded from the scope of protection; the personalized move sequences generated by grandmasters' autonomous tactical reasoning under match pressure meet the above minimal originality threshold.

The judgment of Carlsen v. Chess AI Tech Inc., 321 F. Supp. 3d (S.D.N.Y. 2024) further consolidates the above jurisprudential logic [1]. The district court held that each complete two-player official chess match constitutes a joint collaborative work created by both competing grandmasters.Pre-set opening databases and mechanical computational heuristics can never reproduce the context-sensitive dynamic decision-making track of real matches, which fully meets the two statutory prerequisites of fixation and originality required by U.S. copyright law. The court also specially pointed out that competitive chess is different from purely objective data statistics; each game contains abundant subjective strategic trade-offs of players, so unauthorized mass scraping and commercial TDM of complete game records shall be identified as copyright infringement. Up to now, no U.S. federal court has held that the bulk extraction of grandmaster game records for paid AI training belongs to legitimate fair use behavior.

DOI: 10.54254/2753-7048/2026.35831

2.3. Jurisprudential foundation of chess game copyright protection under China's Copyright Law

The originality evaluation standard specified in Article 3 of the revised 2021 Copyright Law of the People's Republic of China is logically consistent with the U.S. idea-expression dichotomy under the framework of civil law. Fixed unified chess framework rules are shared public resources without proprietary exclusivity, while the unique tactical choices formed in the confrontation between two grandmasters condense independent intellectual input and deserve complete property right protection. In recent years, Chinese people's courts have issued multiple typical digital sports copyright judgments, which continuously confirm that competitive sports records with independent personalized decision-making logic are original intellectual creations protected by copyright law.

The multi-layered overall layout, real-time adaptive tactical adjustment and risk weighing thinking embedded in each chess game cannot be mechanically replicated by unified computer algorithms, fully complying with all statutory constitutive elements of copyrightable works in China's legislation. Different from rigid, purely objective sports statistical indicators such as winning and losing data, high-level chess game notations carry rich differentiated expressive content generated by human subjective judgment, and cannot be simply classified as unprotected factual information. Once an AI enterprise crawls thousands of game records without authorization and uses them for paid commercial products, grandmasters have the right to file civil lawsuits to claim cessation of infringement, compensation for losses and elimination of adverse impacts in accordance with Chinese copyright judicial practice. Relevant domestic academic research has verified the copyright originality of chess competitive notations and delivered solid theoretical grounds for defining the creative attributes of grandmasters' successive competitive move sequences [2, 3].

3. Comparison of copyright regulatory rules on commercial chess data crawling

3.1. Current copyright regulatory regime of the United States

U.S. courts evaluate the eligibility of commercial chess data TDM for fair use through holistic weighing of the four-factor standard laid out in 17 U.S.C. § 107, and large-scale full-text scraping of grandmaster game notations to fuel paid AI training cannot pass such comprehensive judicial assessment [4].

These scraping operations solely serve to develop subscription-oriented commercial products, lacking transformative commentary, academic research or other derivative creation that satisfies the core requirement of legitimate fair use; each complete chess notation is copied in its entirety rather than excerpted partially. Unregulated mass collection of this kind directly undermines potential authorized licensing markets for grandmasters, and sustained large-scale scraping creates lasting adverse market harm to creators' legitimate commercial interests.

Drawing on Judge Leval's authoritative scholarship on transformative use, courts consistently refuse to recognize purely extractive data mining lacking secondary re-creation as a qualifying fair use act, a position uniformly upheld across all existing federal judicial precedents to date [5].

From the perspective of copyright economics, the analytical paradigm proposed by Xiong Qi distinguishes pure data extraction from transformative derivative creation; scholarly research by Zhu and Jiang further confirms that mere full replication of chess games for AI training fails to satisfy the legal standard of transformative use [6, 7].

Fernandez's preprint systematically analyzes the structural defects of copyright limitation rules governing large-scale commercial text and data mining, which further explains why mass

DOI: 10.54254/2753-7048/2026.35831

unauthorized scraping of chess game records cannot obtain statutory fair use exemptions [8].

In terms of infringement remedy rules, the damage compensation clause of 17 U.S.C. § 504 is designed for single independent works, and there are no supplementary provisions aiming at mass database infringement composed of tens of thousands of scattered game records. Even if commercial chess AI platforms obtain millions of dollars in annual operating income relying entirely on unlicensed crawled game corpora, judges can only calculate compensation based on a single game as the basic unit, resulting in nominal damage awards far lower than the grandmasters' actual licensing losses and litigation costs, which cannot form effective infringement deterrence. Meanwhile, § 106 of the U.S. Copyright Act stipulates that all commercial exploitation behaviors must obtain separate written authorization from each copyright owner. This one-to-one licensing mode is suitable for small-volume single content transactions, but when applied to cross-border databases integrating game records of tens of thousands of grandmasters from multiple countries, it will bring huge time, labor and legal compliance costs to AI operators, forming an insurmountable institutional barrier to legitimate data acquisition.

In addition, the U.S. legal system still lacks binding circuit-level special precedents and unified industry self-regulatory guidelines for automated large-scale chess data mining. Mainstream global database operators represented by ChessBase formulate unilateral user service agreements, which unilaterally obtain permanent worldwide royalty-free commercial exploitation rights of game records uploaded by users, without any mandatory profit-sharing obligation for the original creating grandmasters, further magnifying the imbalance of industrial interest distribution.

3.2. Current copyright regulatory regime of China

Different from the open balancing standard adopted by the United States, Article 24 of China's 2021 Copyright Law sets a closed enumeration catalogue of fair use scenarios, with extremely limited space for judicial expanded interpretation [9]. Statutory exemptions only cover non-profit personal learning and limited copying for classroom teaching, and completely exclude profit-driven TDM behaviors for commercial chess AI training. Kong proposes diversified governance solutions to balance copyright protection and data utilization in online environments [10]. In combination with the three-step test followed by all Chinese copyright judicial practices, any secondary use that impairs creators' normal commercial market operation rights shall be identified as infringement. Large-scale batch crawling of complete tournament game notations deprives grandmasters of potential licensing revenue, which clearly violates the statutory provisions of China's Copyright Law and shall bear corresponding civil liability.

In the aspect of liability setting, Article 54 of China's Copyright Law stipulates a unified maximum compensation cap of 5 million RMB per single lawsuit, without tiered punishment standards linked to the quantity of misappropriated independent game records. This flat penalty mechanism leads obvious unfairness in liability allocation: enterprises that illegally crawl hundreds of thousands of professional match notations bear almost the same legal consequences as individual users who copy a small number of games for personal viewing, and the deterrent effect on large-scale database infringement is extremely weak. Similar to the licensing dilemma in the United States, Article 10 of China's Copyright Law requires separate consent of each copyright owner for all commercial utilization acts. Professional grandmasters mainly rely on tournament bonuses as their main source of income, and there is no special intellectual property team to conduct repeated scattered licensing negotiations with countless technology enterprises every year. The separation of decentralized individual copyright ownership and centralized mass data demand pushes almost all commercial chess data collection behaviors into unregulated legal gray zones.

DOI: 10.54254/2753-7048/2026.35831

Furthermore, the Supreme People's Court of China has not issued special judicial interpretations or guiding cases targeting mass scraping of sports competition records for paid AI training, and national copyright administrative authorities have not released targeted industrial supervision documents for the commercialization of chess digital databases [11]. The normative boundary between non-commercial academic research and profit-driven commercial data extraction is ambiguous, which further aggravates the chaos of the whole industry's data utilization order.

3.3. Comparative analysis of shared institutional defects

Although China and the United States maintain consistent judicial consensus on the copyright originality of chess game records, the two copyright systems share four interrelated structural institutional defects, which together lead to long-term unbalanced profit distribution in the global chess AI industry. To begin with, neither country's fair use framework can provide legal exemption channels for large-scale commercial TDM and bulk crawling of chess competition data, making it difficult for AI enterprises to carry out compliant data collection through low-cost means. Meanwhile, the damage compensation systems of both countries are formulated aiming at isolated single-work infringement cases, and cannot adapt to the unique characteristics of mass database infringement composed of a large number of scattered independent game records, resulting in insufficient judicial economic remedies for infringed grandmasters. Moreover, the statutory requirement of one-to-one individual licensing brings prohibitive cross-border transaction costs to all chess data operators, which fully reflects the practical necessity of building a unified collective licensing mechanism. Last but not least, neither jurisdiction has formulated targeted special regulatory norms for sports competition data used as AI training materials, leaving long-standing legislative loopholes that hinder unified cross-border copyright governance of chess data [12]. Under the superposition of these four interrelated institutional deficiencies, almost all commercial value generated by grandmasters' years of intellectual labor is occupied by AI technical enterprises, while professional athletes lack feasible legal channels to recover economic losses through civil litigation, forming a long-term unfair industrial interest pattern.

4. The development direction of copyright governance rules for commercial chess AI game databases

To eliminate the four homologous institutional defects summarized through comparative legal analysis, this paper constructs an innovative transnational dual-track governance framework integrating coordinated legislative revision and unified global sports industry self-regulation, so as to balance the sustainable commercial development of chess AI and full proprietary copyright protection of professional grandmasters. The framework is divided into legislative optimization and industry self-regulation two parallel implementation paths, with complementary functions and mutual supervision mechanisms.

4.1. Legislative reform track: establish tiered differentiated mass infringement compensation standards

For the United States, it is suggested that legislative authorities supplement new subsections to 17 U.S.C. § 504. Two quantifiable metrics--the volume of unique original chess records scraped without permission and the annual revenue generated by AI tools trained on such unlicensed data--may serve as core benchmarks for calculating aggregate infringement compensation. Policymakers

DOI: 10.54254/2753-7048/2026.35831

may set a revenue threshold for small AI startups, allowing them to adopt lower per-record compensation benchmarks and easing compliance burdens for emerging small and medium-sized developers. Meanwhile, profitable large tech platforms that engage in intentional mass unauthorized crawling may face multiplied punitive damages tied to their illegal operating gains, which can substantially enhance civil liability deterrence against large-scale database infringement.

For China's copyright law revision, national legislators may consider amending Article 54 of the 2021 Copyright Law. Abolishing the uniform single-case compensation ceiling of 5 million RMB and establishing damage calculation standards tied to the actual volume of infringed game notations can eliminate unfair flat penalty rules that fail to distinguish infringement scale and undercompensate grandmasters for their economic losses. Coordinated legislative adjustments in China and the U.S. can align judicial sanctions with athletes' real market losses and enterprises' illegal profits, fulfilling copyright law's dual purposes: compensating creators and deterring infringing conduct.

4.2. Industry self-regulation track: fide-managed global unified collective licensing system

As the highest authoritative international organization governing global chess events, the International Chess Federation (FIDE) is advised to set up a permanent special international chess copyright foundation to be fully responsible for centralized standardized collective licensing of all global commercial chess databases and AI training platforms, thoroughly replacing the inefficient and operationally unworkable one-to-one individual authorization model [13]. This foundation would first develop a universal cross-border framework licensing template applicable to all chess data operators. After finalizing unified licensing compliance terms, it can sign formal cooperation agreements with mainstream commercial chess AI platforms and database suppliers, collecting a fixed proportion of royalty revenue from enterprises' annual data-driven business gains. All collected licensing proceeds will be allocated proportionally to professional grandmasters based on the volume of their original game records contained within each commercial training dataset.

This centralized collective licensing mechanism can drastically cut redundant cross-border negotiation, audit and filing transaction costs for AI developers, while providing stable long-term recurring royalty income for chess creators as legal consideration for their proprietary copyright exploitation rights [14, 15]. It is proposed that China's National Copyright Administration and the U.S. Copyright Office conduct joint cross-border regulatory oversight and craft globally consistent licensing compliance criteria to address parallel regulatory loopholes within both national legal frameworks at the same time. Relying on FIDE's inherent global sports management authority, this mechanism can realize standardized cross-border data authorization without large-scale comprehensive revision of national copyright statutes of various countries, with strong operability and promotion value.

4.3. Generalized reference value for global competitive sports industry

The dual-track cross-border governance framework constructed in this paper is not limited to the chess industry alone. The cross-border governance framework constructed in this paper can be generalized to various competitive sectors in accordance with the unified sports intellectual property governance logic summarized in WIPO's industry reference guide [16]. After targeted parameter adjustment according to the event characteristics of different sports, the whole set of institutional design can be transplanted to resolve similar TDM-related copyright disputes triggered by training data of basketball, football, table tennis, e-sports and other competitive sports. The tiered

DOI: 10.54254/2753-7048/2026.35831

differentiated compensation rule and FIDE-style unified collective licensing model jointly form a balanced adjustable regulatory template, which can coordinate two major public interests worldwide: the innovative growth of sports artificial intelligence industry, and comprehensive proprietary copyright protection of athletes who create core competitive data assets. This research provides a replicable cross-border governance solution for all sports AI data copyright conflicts, with broad theoretical and practical significance.

5. Conclusion

Either extreme regulatory idea of completely deregulating unlicensed chess data crawling or comprehensively prohibiting all commercial extraction of game records cannot resolve the fundamental structural contradiction between the expansion of chess AI industry and grandmasters' copyright protection. The four homologous institutional defects shared by China and the U.S. copyright systems have long sustained unregulated legal gray zones in the global chess data industry and formed severely distorted interest distribution relations between technical enterprises and athlete creators. The dual-track cross-border governance framework combining targeted legislative revision and FIDE-led global collective licensing proposed in this paper sets tiered differentiated compensation standards to correct unfair flat liability rules for mass database infringement, and replaces costly individual authorization with centralized collective licensing to reduce redundant industrial transaction costs. This institutional design conforms to the core legislative logic of copyright law in both China and the United States, realizes long-term balanced and equitable benefit distribution between AI enterprises and professional chess players, and provides replicable cross-border governance experience for copyright disputes of sports AI training data all over the world.

References

[1] Judicial case. (2024). *Carlsen v. Chess AI Tech Inc., 321 F. Supp. 3d* (S.D.N.Y. District Court).
[2] Xu, W. K. (2021). Legal protection paths for tournament data rights in the digital sports era. *Journal of Sport Science, 41*(7), 79–87.
[3] Lele, W. (2026). Copyright protection of competitive chess game notations. *International Intellectual Property Law Review, 38*, 112–128.
[4] Sag, M. (2023). Copyright safe harbors for algorithmic text and data mining. *Journal of the Copyright Society of the USA, 70*(2), 189–221.
[5] Tyagi, K. (2024). Copyright and text & data mining for generative AI training. *Journal of Intellectual Property Law & Practice, 19*(7), 557–570.
[6] Xiong, Q. (2024). An economic analysis paradigm of copyright law for secondary creation and text & data mining. *Contemporary Law Review, 1*, 108–120.
[7] Zhu, W. Y., & Jiang, B. B. (2022). The copyright boundary of data crawling under transformative use. *Academic Exchange, 3*, 55–66.
[8] Fernandez, R. (2025). Copyright limitations for commercial text and data mining. *Johns Hopkins University Preprint*.
[9] Tang, S. H. (2018). *Research on copyright exception system for large-scale text and data mining*. Retrieved from http://iolaw.cssn.cn/fxyjdt/201802/t20180216_4656899.shtml.
[10] Kong, X. J. (2020). Optimizing copyright rules in the internet era: From notice-and-takedown to multi-dimensional governance. *Journal of Political Science and Law, 1*, 52–66.
[11] Feng, G. (2026). Copyright qualification and judicial protection path of competitive sports data. *Pengpai Law Review, 10*, 92-101.
[12] Prokopiev, I. (2026). *AI training data copyright: U.S. vs. EU comparative analysis*. Retrieved from https://www.licentium.io/post/ai-training-data-copyright-and-transparency-a-comparative-legal-analysis.
[13] FIDE. (2025). *Official media & copyright management guidelines for world chess tournaments*. Retrieved from https://www.fide.com/docs/candidates2022/Photography%20and%20Videography%20Guidelines.pdf.

DOI: 10.54254/2753-7048/2026.35831

[14] Wang, W. Q. (2026). Construction of legal licensing mechanism for sports training data under machine learning scenarios. *East China University of Political Science and Law Journal, 2*, 134-147.
[15] Xiong, J. (2025). Optimization path of collective copyright licensing system adapted to AI data industry. *Intellectual Property, 8*, 45-54.
[16] WIPO. (2022). *Reference guide to intellectual property rights in global sports industries*. Geneva: World Intellectual Property Organization.