A Study on Big Data Legal Supervision

Authors: Liu Long

Date: 2024-01-23T00:00:00+00:00

Abstract

Big data legal supervision is a current hot topic and the key to the new round of prosecutorial system reform. It poses new requirements for procuratorial organs and necessitates a grasp of its fundamental theories. As an emerging phenomenon, big data legal supervision is still in a phase of vigorous development, inevitably encountering various problems, among which the most prominent and controversial are the three aspects of data, algorithms, and institutions. For big data legal supervision to fulfill its intended role, improvements are needed from multiple perspectives. This paper addresses the existing problems and, based on China's actual conditions, proposes several improvement measures to provide some reference for the formulation of relevant regulations.

Full Text

Research on Big Data Legal Supervision

Big data legal supervision represents both a contemporary hot topic and a cornerstone of the current round of prosecutorial system reform. It imposes new requirements on prosecutorial organs and necessitates a solid grasp of its fundamental theories. As an emerging phenomenon still in vigorous development, big data legal supervision inevitably encounters various challenges, with the most prominent and contentious issues concentrated in three areas: data, algorithms, and institutions. To enable big data legal supervision to fulfill its intended functions, coordinated improvements from multiple perspectives are essential. This paper identifies existing problems and proposes several measures for improvement based on China's actual conditions, offering some reference for the formulation of relevant regulations.

Keywords: big data, legal supervision, algorithm, institution

1. Introduction

General Secretary Xi Jinping has pointed out that "practice has repeatedly taught us that core technologies in key fields cannot be begged for, bought, or bargained for," that "for China to become strong and achieve rejuvenation, it must vigorously develop science and technology," that "we must seize the opportunities presented by the integrated development of digitalization, networking, and intelligence, using informatization and intelligence as levers to cultivate new drivers of growth," and that we must "use 'innovation' to drive 'reform' and use increments to drive the existing stock." As an important component of China's state apparatus, prosecutorial organs bear the vital functions of legal supervision and crime punishment. On the one hand, they should improve the rule-of-law environment for protecting innovation and provide more comprehensive judicial guarantees for scientific and technological innovation; on the other hand, they need to deepen the construction of smart prosecutorial work, keep pace with the times, seize opportunities, confront problems head-on, and rise to challenges shouldered by history.

On June 15, 2021, the Central Committee of the Communist Party of China issued the *Opinions on Strengthening Legal Supervision Work by Prosecutorial Organs in the New Era*, explicitly proposing requirements for prosecutorial organs to give full play to their legal supervision functions and strengthen intelligent construction. In January 2022, Procurator-General Zhang Jun emphasized at the National Chief Procurators (Enlarged) Conference the need to enhance strategic thinking on big data and use big data to help achieve fundamental improvements and qualitative transformations in legal supervision. The proposal of the prosecutorial big data strategy represents an important measure for prosecutorial organs to thoroughly study and implement Xi Jinping Thought on the Rule of Law and General Secretary Xi Jinping's important discussions on building a strong cyber nation and digital China, to implement new development concepts, and to adapt to new trends in the information age. It fully demonstrates the sense of responsibility and mission of prosecutorial organs to meet the needs of modern social governance, helps strengthen legal supervision methods, enhances the effectiveness of legal supervision, accelerates quality, efficiency, and power transformations in prosecutorial work, and boosts the modernization of the national governance system and governance capabilities.

Implementing the prosecutorial big data strategy requires genuine and profound efforts in strengthening ideological understanding, expanding information sources, enhancing business training, and promoting comprehensive applications, thereby contributing greater prosecutorial strength to promoting high-quality development of prosecutorial work and advancing the modernization of the national governance system and governance capabilities. With the rapid development of information technology and the in-depth implementation of the national big data strategy, big data is playing an increasingly important role in the high-quality development of the economy and society. Advancing the rule of law in all respects and promoting the modernization of the national gover-

nance system and governance capabilities cannot be achieved without big data. General Secretary Xi Jinping has repeatedly stressed the need to use big data to enhance the modernization of national governance. Regarding big data empowerment of legal supervision, Procurator-General Zhang Jun offered a vivid analogy: "Supervision and case handling are like farming. While one cannot do without traditional knowledge and basic agricultural techniques for spring planting, summer weeding, and autumn harvesting, in the information age, having only these traditional, basic knowledge and skills is far from sufficient. We must use technology and big data methods to improve quality and efficiency to increase 'productivity'!" How to enhance prosecutors' ability to deeply apply big data to achieve precise supervision? How to use big data to drive qualitative transformation of legal supervision on the basis of fundamental improvement? How to promote solutions to deep-seated social governance problems? These have become unavoidable questions before legal professionals.

2. Basic Theory of Big Data Legal Supervision

2.1 Conceptual Analysis A fundamental task of theory formation is to reasonably define the concept of a particular phenomenon. As a new phenomenon, the connotation and extension of big data legal supervision are not theoretically very clear, making conceptual analysis indispensable. However, many prosecutorial organs in practice remain confused: some conflate it with auxiliary case-handling tools such as complaint inquiry, fund flow analysis, and electronic evidence investigation, which are mainly used for collecting clues and are not entirely consistent with big data legal supervision; others confuse it with post-facto data statistics for thematic research or trend prediction, which merely record data and cannot discover unknown clues from data; still others conflate it with digital case management, such as discovering case omissions or errors through electronic file sharing and case data analysis, which only focuses on technical convenience and efficiency rather than the big data legal supervision that requires key reform.

In simple terms, big data legal supervision should be legal supervision driven by big data, with big data injecting new impetus into legal supervision. The use of big data actually dates back to the early days of the People's Republic, when staff evaluated annual weather data to select National Day and found that October 1 had the most sunny days in historical data, thus choosing it as the date. When applying big data to legal supervision, it can be simply expressed as "legal supervision + big data." Here, legal supervision refers to "the specialized work of specific state organs, authorized by law and using legally prescribed means to monitor and supervise law implementation with statutory effect," representing the special supervision exercised by China's prosecutorial organs according to their constitutional and legal duties. Big data can be simply understood as "the collective term for large datasets, big data technologies, and big data applications that serve decision-making problems." Additionally, it should be noted that, unlike foreign countries, relevant explorations of big data

legal supervision in China present a special pattern of being "dominated by state power."

2.2 Basic Characteristics Big data legal supervision discovers patterns and problems from data or individual cases, develops supervision models through summarization and induction, mines supervisory information and class-action clues hidden in massive data that are difficult to know based on individual cases alone, and then analyzes systemic loopholes in management connections, institutional mechanisms, and legislation-enforcement-judiciary from class-action problems to promote social governance through guiding investigations and proposing prosecutorial suggestions. Specifically, big data legal supervision has four characteristics: First, it proceeds from individual cases to class actions and then to social governance. The path of big data legal supervision involves summarizing patterns and characteristics from individual cases, then screening class actions from massive data based on these patterns, discovering loopholes in legislation, law enforcement, judiciary, and mechanisms from batch class actions, with the ultimate goal of solving social governance problems. Second, it precisely identifies class-action clues. Big data legal supervision needs to start from specific supervision scenarios or small entry points, extract effective class-action characteristics and correct logical patterns, and ultimately accurately identify class-action supervision clues from massive data. Third, supervision targets are universal. The supervision objects or loophole problems targeted by big data legal supervision have broadness and universality, with batch class actions existing in reality and also present in other places or fields, holding social governance significance. Fourth, it is technically feasible. Big data legal supervision must be realistic and feasible technologically to be applied, replicated, and promoted in practice. Data must be collectable, transformable, and cleanable, capable of mining hidden information behind data through specific methods such as screening, collision, clustering, and statistics.

2.3 Significance As a round of reshaping and transformation in legal supervision, big data legal supervision ingeniously constructs an Archimedes lever model. Archimedes discovered that infinite force could be saved with the help of a lever. Specifically, the product of the force and force arm on both ends of the lever is equal, allowing the lever to balance. The formulaic expression of the lever principle is: power $\times$ power arm = resistance $\times$ resistance arm. This principle shows that in rule-of-law construction, the more the power and power arm end is expanded, the more goals can be achieved; similarly, the more the resistance and resistance arm end is compressed, the better the effect. The secret of big data legal supervision boosting rule-of-law construction lies here.

First, big data legal supervision expands the "kinetic energy" for prosecutors to identify illegal cases. Traditionally, legal supervision was initiated from individual case clues through channels such as party reports, with prosecutors launching investigations and verifications limited by single clues and restricted channels. Big data legal supervision has equipped prosecutors with wings of big

data analysis, data models, and platforms, opening up new horizons for class-action supervision and intelligent regulation. For example, the Hubei Provincial People's Procuratorate's government data cloud supervision platform can analyze and compare millions and tens of millions of administrative law enforcement case data across the province, while the Shaoxing City People's Procuratorate in Zhejiang Province's false litigation supervision model can detect false litigation case clues in Beijing and other places, representing unimaginable leaps. Meanwhile, big data legal supervision can "profile" abnormal cases and transform them into intelligent prediction and early warning conditions, achieving the effect of "top-tier doctors treating diseases before they occur."

Second, big data legal supervision extends the "power arm" for prosecutors to govern illegal acts. China's prosecutorial system has long suffered from the phenomenon of supervision and case handling being "two separate skins." Big data legal supervision tightly integrates legal supervision into the "four major prosecutorial functions" and "ten major businesses," achieving benign linkage between business departments, case management departments, and technical departments, and giving rise to new integrated supervision. More importantly, big data legal supervision implements a "one-axis, multiple-elements" co-governance structure, led by prosecutorial organs and jointly involving public security organs, adjudicative organs, judicial administrative organs, administrative organs, as well as social organizations and industry associations, achieving true "win-win and multi-win" through promoting multi-party cooperation.

Third, big data legal supervision compresses the existence space and destructive power of illegal cases. This is mainly manifested in its decomposition of the radiation surface and "circle of friends" of illegal objects, curbing the adverse effects of continued illegal acts. The governance advantages demonstrated are: first, intelligently locking onto various batch abnormal cases for multi-case correction; second, analyzing routine cruxes of social governance for class-action error prevention. For example, the Suzhou City People's Procuratorate in Jiangsu Province's "bankruptcy field false labor-capital creditor's rights supervision model" has screened hundreds and thousands of clues involving labor disputes in risky enterprises, comprehensively correcting all batch illegal cases involving labor remuneration disputes discovered. Subsequently, the procuratorate deepened information sharing mechanisms and built a long-term governance system, generating commendable governance effectiveness.

3. Concerns Regarding Big Data Legal Supervision

Scholars have pointed out that e-government project risks cover various elements of e-government projects, such as project scope risk, project cost risk, project schedule risk, project quality risk, and project team risk, with various project risks being interrelated and superimposed. As an important component of e-government, big data legal supervision is no exception. While recognizing its significance, we should also maintain a clear mind and provide timely prevention against possible risks. The following briefly analyzes the concerns and risks of

big data legal supervision.

"Corresponding to disciplinary society are high-energy machines; corresponding to control society are cybernetics or computers… In the face of those uninterrupted control forms appearing in open environments, perhaps the harshest confinement seems like wonderful memories to us." Deleuze's concerns are not without reason, especially when the development of big data intelligence has exceeded people's imagination, causing more fear than joy. A simple 梳理 from three aspects—data, algorithms, and institutions—reveals several concerns regarding big data legal supervision.

3.1 Data Aspects Data can be said to be the foundation of big data legal supervision, a brick in the Great Wall. However, a dike of a thousand miles often collapses from an ant hole. In terms of data, several risks may exist that, when serious, could affect the fair exercise of prosecutorial power: First, data may be incomplete. Comprehensiveness is one of the basic requirements of big data legal supervision, and the completeness of data volume directly determines the accuracy of analysis results. Mechanisms such as evidence review and sentencing deviation analysis in big data judicial supervision overly rely on data from "China Judgments Online." However, the more than 30 million judgment documents publicly available online actually only account for about 50% of the total, with approximately half of judgment documents not published online. In addition to final judgment documents, the digitization of procedural documents such as investigation conclusion reports and prosecution opinions is also insufficient. Second, data may be untrue. Data itself may contain errors, and drawing wrong conclusions based on erroneous data is unavoidable. Specifically regarding big data legal supervision, ensuring correct identification of much data under judgments mixed with many subjective factors has become a top priority. For instance, different determinations made by different judgments for similar behaviors—such deviations and erroneous judgment data create "data noise," equivalent to using wrong judicial experience and adjudication rules to guide individual cases, inevitably causing algorithm models based thereon to contain certain biases. Third, data may be unreliable. This point is actually a superimposed state of the previous two, further exacerbating the impact of data on big data legal supervision. It is not difficult to imagine that in this era of data explosion, new data is generated daily at a dizzying pace. For big data legal supervision, the judicial field produces massive new judgment documents every day, and the continuous influx of new data means it may contain new adjudication rules and supervision rules. New judicial data brings new adjudication experience and supervision rules, causing existing judicial supervision models to be insufficiently comprehensive and refined. Moreover, as big data technology develops rapidly, existing judicial supervision models will increasingly lag behind new judicial power operation practices.

3.2 Algorithm Aspects If data is the brick of big data legal supervision, then algorithms are the method for building the Great Wall—having raw materials

alone without a manufacturing process will not work. In terms of algorithms, several risks may exist that have been continuously discussed in academia: First, algorithms may not be intelligent. Algorithm generation in big data judicial supervision mainly has two approaches: "bottom-up" and "top-down." The former refers to using certain technical means to extract resource patterns from massive data, with machines automatically generating algorithm models after learning from massive data; the latter refers to using established judicial rules and experience to artificially construct algorithm rules. Currently, given the limitations of artificial intelligence development levels and the inherent difficulty of the judicial profession, most algorithm generation still adopts the "top-down" approach, translating subjective judicial experience into algorithmic rules. The so-called "how much human labor, how much intelligence" indicates that the algorithm generation process in big data judicial supervision remains insufficiently automated and intelligent. The real legal world contains large amounts of legal logic not yet infiltrated by experience, as well as experience not yet regulated by logic. This may lead to problems such as incomplete crime coverage and imprecise correspondence. Second, algorithms may be unjust. Whether for supervisors or those being supervised, algorithms are actually opaque, often creating a so-called "black box effect." While bringing convenience to prosecutors and saving learning costs, we should 清醒地认识到 that algorithm non-disclosure is highly likely to cast a veil of injustice over big data legal supervision, reducing public trust, and even technical issues such as what data will be collected and whether parameters are correct may be difficult for prosecutors to verify. This will lead to opacity in the judicial supervision process, thereby causing procedural injustice in judicial supervision. Third, algorithms may be differentiated. This seems to be a long-standing concern in the industry, with "big data price discrimination" exposed at the 315 Gala still fresh in memory. China has a vast territory with inherent differences between regions, so how standards are set and whether differentiated algorithms will further lead to unequal treatment of cases warrants further research.

3.3 Institutional Aspects Relatively speaking, institutional risks are more macroscopic, concerning conceptual issues. First, any power carries the risk of abuse, and big data legal supervision should have limits. Abuse may lead to disorderly expansion of prosecutorial power, affect the enthusiasm of those being supervised, create fear and resistance, and even lead to countermeasures to evade big data legal supervision. Over time, this will compress the operational scope of judicial power and rigidify its operation. Second, mechanical adjudication leads to 固化 thinking, with courts across the country increasingly becoming "judgment factories," creating scenarios where algorithms dominate individual case justice. Faced with legal disputes, judicial personnel no longer resort to strict adversarial trials, legal reasoning, and other procedural requirements, but instead resort to machine judgment results. The essence of judicial adjudication is value judgment, while the essence of algorithms is data operation, which cannot make value judgments with emotional coloring like judicial officers.

4. Countermeasures for Related Concerns

Through the above analysis, it is not difficult to find that concerns about big data legal supervision have four characteristics: First, low tolerance—once major problems occur, the consequences are immeasurable; second, high probability of sudden occurrence—involving more than 3,000 prosecutorial organs at four levels nationwide and hundreds of thousands of staff; third, long control process—requiring a stable, high-quality work team; fourth, diverse element categories—must run through all stages.

First, strengthen the Party's leadership over big data legal supervision. Precisely based on this characteristic of China's political system, central Party and government departments issue unified standards and regulations to top-down promote prosecutorial supervision intelligence and its specific requirements, which will make the implementation of big data legal supervision reform feasible.

Second, strengthen data usage in big data legal supervision. Enhance data collection, analysis, and judgment, achieve interconnectivity and information sharing, legally and forcefully supervise and guide judicial organs to comprehensively obtain electronic data and other objective evidence, form complete evidence chains, lay a solid foundation for smooth case handling, and improve the quality of precise case handling.

Third, strengthen algorithmic control over big data legal supervision. The big data legal supervision mechanism should have verifiability and transparency. Data sources and algorithm principles should be disclosed to supervised objects to safeguard procedural legitimacy in judicial supervision. In fact, algorithm explainability and even visualization are technically solvable problems. Once the law puts forward relevant requirements, the technical community will find ways to make algorithms explainable.

Fourth, strengthen institutional design for big data legal supervision. Clarify the responsibility chain of big data legal supervision, clearly distinguish responsibilities between different subjects, such as the responsibilities of machines and technical personnel, system deployers, and judicial officers, and implement a series of requirements through clear institutions. Fully and reasonably utilize digital technologies such as artificial intelligence and big data, comply with and leverage the trend of judicial digitalization, and promote the construction and improvement of China's judicial responsibility system.

Fifth, establish and improve the evaluation mechanism for big data legal supervision. Every year, political and legal committees at all levels, together with judicial organs, should conduct comprehensive and objective evaluations of big data legal supervision at their respective levels, publish evaluation reports, supervise rectification, smooth feedback channels, and actively listen to public voices.

5. Conclusion

British scholar Richard Susskind in his book *The Future of the Professions* warns legal professionals to think wisely about the future. The future has arrived and is approaching; having seen the future, why not prepare? Compared with understanding that remains in the past and present, predicting the direction of the human world is a broader and more meaningful challenge. For the exploration of big data legal supervision, this requires the prosecutorial community to be courageous actors, achieve the digital prosecutorial transformation, and rely on big data technology to launch multi-case supervision. For advancing national and social governance modernization, prosecutorial organs and relevant agencies should expand the "concentric circles" and collaboratively build "China's governance" through the win-win and multi-win approach of legal supervision. The entire process of institutions gradually accumulates in the long river of years passed and precipitates through the historical push of the people. Big data legal supervision is still in vigorous development, requiring rational and dialectical views on the formation and evolution of new institutions, and also leaving them to be tested by practice and time.

References

[1] Zhao Zhigang, Jin Honghao. *Introduction to Smart Prosecutorial Work: The Technological Wisdom of Prosecutorial Legal Supervision*. China Procuratorial Press, 2018.

[2] [France] Gilles Deleuze. *Negotiations: Interviews with Gilles Deleuze*. Translated by Liu Hanquan. The Commercial Press, 2000.

[3] [UK] Richard Susskind. *Tomorrow's Lawyers: An Introduction to Your Future*. Translated by He Guangyue. Peking University Press, 2015.

[1] Weng Yueqiang, Wen Xiaoqing, Jiang Qi, et al. The Value and Application of Big Data Empowering Legal Supervision[J]. *People's Procuratorate*, 2022(11):27-30.

[2] Zhang Zhihui. Three Discriminations of Legal Supervision[J]. *China Legal Science*, 2003(05):14-22. DOI:10.14111/j.cnki.zgfx.2003.05.002.

[3] Zhu Yangyong, Xiong Yun. Is Big Data Data, Technology, or Application?[J]. *Big Data*, 2015,1(01):78-88.

[4] Liu Pinxin. China's Innovation in Smart Justice[J]. *Journal of National Prosecutors College*, 2021,29(03):81-101.

[5] Weng Yueqiang, Wen Xiaoqing, Jiang Qi, et al. The Value and Application of Big Data Empowering Legal Supervision[J]. *People's Procuratorate*, 2022(11):27-30.

[6] Zuo Weimin. Some Thoughts on the Application Prospects of Legal Artificial Intelligence in China[J]. *Tsinghua Law Journal*, 2018,12(02):108-124.

[7] Liu Qiao, Li Yang, Duan Hong, et al. A Survey on Knowledge Graph Construction Techniques[J]. *Journal of Computer Research and Development*, 2016,53(03):582-600.

[8] Zheng Ge. The Law of Algorithms and the Algorithms of Law[J]. *China Law Review*, 2018(02):66-85.

[9] Ji Weidong. Changes in Judicial Power in the Age of Artificial Intelligence[J]. *Oriental Law*, 2018(01):125-133. DOI:10.19404/j.cnki.dffx.2018.01.013.

[10] Wang Ran. Research on the Mechanism of Big Data Judicial Supervision[J]. *Journal of Hunan University of Science and Technology (Social Science Edition)*, 2021,24(03):132-141. DOI:10.13582/j.cnki.1672-7835.2021.03.017.

[11] Zhao Yiyu. Development Concerns and Response Logic of Intelligent Prosecutorial Supervision[J]. *Law and Social Development*, 2023,29(02):178-194.

[12] Zhang Junjie, Ma Xiaoyi, Liu Deyin. Discussion on the Innovative Mechanism of Big Data + Prosecutorial Supervision and Management Operation[J]. *Chinese Prosecutor*, 2018(19):25-28.

[13] Zheng Ge. The Law of Algorithms and the Algorithms of Law[J]. *China Law Review*, 2018(02):66-85.

[14] Gao Tongfei. Upholding Judicial Responsibility Ethics in the Digital Age[J]. *Law and Social Development*, 2022,28(01):151-172.

[15] Liu Pinxin. On Big Data Legal Supervision[J]. *Journal of National Prosecutors College*, 2023,31(01):76-92.

[16] Zhu Suli. How Are Institutions Formed?—The Story of Marbury v. Madison[J]. *Comparative Law Study*, 1998(01):68-74.

Xi Jinping's speech at the 19th Academician Conference of the Chinese Academy of Sciences and the 14th Academician Conference of the Chinese Academy of Engineering.

Note: Figure translations are in progress. See original paper for figures.