The Intervention, Influence, and Disruption of Big Data in the Legal Domain

Authors: Yingyi Chen

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

Abstract

none

Full Text

Introduction

The book *Data-driven Law: Data Analytics and the New Legal Services*, which Professor Zhou introduced to us in class, presents and examines the various interactions between data and law in the legal world. As a typical interdisciplinary study of technology and law, both the book and the course have significantly broadened this author's academic and professional horizons as a law student. The legal world cannot exist in isolation, aloof from society; it must necessarily interact and connect with various factors in the real world. In today's era of rapid development of new technologies such as digitization and big data, data has become an inevitable and long-standing frequent visitor at the gates of law. As data gradually intervenes in this conservative and insular legal world, how will the law embrace it? What role and status will data acquire in the legal world? And what value will data play in the legal world?

The question of whether law should remain relatively closed, adhering to its own distinctive discourse mode, or should appropriately open itself to influences from other disciplines has been a recurring choice in legal research. Examples include economics, sociology, and political science, among others. This modest reading report attempts to further explore this relationship based on the author's understanding of data gained from this book.

What is Legal Big Data?

What is big data? From *Big Data Era: Life, Work, and the Great Transformation of Thinking* by a British scholar, we can roughly understand what big data

entails. So-called big data has three characteristics: full-sample, messiness, and correlation. Among these, the most important is full-sample. Anyone engaged in empirical research knows that when the sample equals the population, sampling error becomes zero. However, due to limitations in financial resources, manpower, and analytical technology, obtaining full-sample data has been difficult. Initially, to understand taxpayers' actual conditions, the state developed various statistical techniques to reduce and control sampling errors. Now, with the development of computer technology, people have discovered that obtaining full-sample data for certain phenomena is not entirely impossible even when facing massive amounts of information. Based on such full-sample data, people can better understand various realities in society, making predictions about the probability of certain phenomena more reliable. Thus, big data is not about the absolute size of the sample, but rather about its "completeness."

Regarding "what is legal big data," the legal community currently has no unified or clear definition. Applying Viktor's understanding of big data, we might conceptualize legal big data as: using an unprecedented approach to analyze massive legal data, predict legal issues, obtain products and services of great value, or derive new understandings, profound viewpoints, and claims. Legal big data may transform the legal services market and organizational frameworks, and even change the relationship between government and citizens. Based on discussions and exchanges with Professor Zhou in class, legal big data does not have a clear boundary defining its content or scope. The concept of legal big data may ultimately become a new way of thinking for legal professionals—a methodology for understanding and analyzing law by combining data analysis to gain empirical insights.

The Intervention of Big Data in the Legal World

Impact on Legal Research

The emergence of legal big data has made possible judicial practices and new forms of empirical research based on legal big data, potentially bringing revolutionary changes to legal research methods. This possibility stems from big data's unique advantages: its "full-sample" characteristic. Big data is typically comprehensive data for a specific field, characterized by massive volume and comprehensive content. Empirical research based on full-sample data can significantly reduce errors that traditional sampling methods may cause, enhance overall understanding of research subjects, and discover information that is difficult or impossible to obtain from traditional sampled data, fundamentally transforming research perspectives, materials, and methods. Additionally, the speed of data generation, collection, and analysis is increasing. "Data analysis is becoming faster and faster, often showing real-time results as soon as data is entered," which helps researchers grasp the full picture of relevant legal practice conditions in a timely and effective manner, overcoming the time-consuming and lagging defects of traditional empirical research methods. Finally, the objectivity and scientific nature of data collection and analysis technologies enhance

research quality.

Unlike traditional artisanal empirical research characterized by personal involvement—where researchers "mostly collect and organize data themselves" and "selectively collect and use data due to research motivations"—massive materials and data are far beyond what researchers in the "manual workshop era" can personally and individually review, count, and analyze. Big data collection and analysis often rely directly on automatic processing and completion by data technology. Under open-source conditions, the research process has considerable transparency, and research conclusions can be reproduced and verified, significantly enhancing the objectivity and scientific nature of data collection and analysis. In particular, datasets collected from different channels generate massive amounts of data. When aggregated, these can be mined and subjected to deeper analysis, revealing various patterns and correlations and enabling statistically meaningful predictions. This not only facilitates diachronic and evolutionary studies [1] but also enables predictive research and trend analysis, ultimately promoting the scientific quality of research.

In recent years, domestic explorations using large amounts of data for legal research have emerged, with scholars already noting ethical issues facing legal big data. Some scholars have offered insightful perspectives on how to conduct big data legal research. However, overall, domestic big data legal research remains in an exploratory stage, with some studies lacking basic understanding of legal big data and their research methods and processes actually built on certain misconceptions. Therefore, examining the current state of big data legal research and clarifying several misconceptions is fundamentally significant for the healthy development of this field.

Impact on Legal Practice

Numerous studies have pointed out that big data analytics has positive significance in promoting national governance decision-making and improving governance capabilities, particularly in providing public services. Regarding legislative activities themselves, big data can also provide better evidentiary support and legitimacy justification, especially in regulatory legislation fields (such as food and drug, production, and environmental safety—data-intensive domains). The traditional legislative drafting process involves investigation, hearings, deliberation, opinion solicitation, and risk assessment, all aimed at obtaining as many ideas and opinions from stakeholders as possible. Analysis based on massive data can provide legislative drafters with more accurate firsthand data, thereby avoiding distortion or neglect of legislative goals by stakeholders acting in their own interests, becoming a powerful supplement to scientific decision-making.

Big data analytics may transcend the de facto boundaries between government and social/private spheres originally caused by insufficient state capacity, thereby requiring legal redefinition of these boundaries and limiting certain an-

alytical and predictive uses. In this sense, technical precision cannot completely replace our pursuit of legislative principles and purposes, and it is necessary to evaluate the expansion of power boundaries and consequences brought by precise technology. Of course, at least for now, due to interference from population mobility and rapid urban development, under the broad framework of using information technology to strengthen social security prevention and control systems, police in some cities can only make rough predictions based on big data. For example, the Beijing Huairou District Police have established a crime data analysis and trend prediction system by integrating historical case information using big data, cloud computing, and scientific analysis models, which can automatically predict crime trends and guide police resource allocation. In addition to eight categories of crimes including burglary, fraud, and robbery, the police have also expanded the system's information input scope to include public security cases, traffic accidents, fire incidents, and other public management events—far from the fine-grained governance targeting individuals [2]. However, it is foreseeable that as databases expand and algorithms improve, big data will play an increasingly important role in future smart city governance and risk prevention.

Currently, China's legal practices using big data are burgeoning. Examples include: vigorously promoting the online publication of judgment documents based on judicial transparency; establishing crime information judgment and trend prediction relying on big data technology; constructing "procuratorial big data standard systems, application systems, management systems, and technology support systems" using big data; establishing case weighting coefficients and evaluation index systems using big data to determine judges' workload and conduct scientific quota allocation and case distribution; and conducting various legal AI practices based on big data, attempting applications such as similar case recommendation, sentencing assistance, and deviation warnings. Among these, the large-scale online publication of judgment documents has given China, for the first time, national, public, and detailed legal data. However, overall, the practical application of legal big data in China remains relatively limited and not widespread, presenting to some extent a phenomenon of "hot discourse, cold practice": on the one hand, the scope of application subjects is limited, mainly concentrated in a few judicial organs and legal data companies; on the other hand, the application fields are relatively narrow and actual applications are few, mainly concentrated in auxiliary case handling aspects such as similar case retrieval, legal document drafting, and intelligent document error correction.

Impact and Challenges of Big Data on the Legal World

Data Security and Information Protection

In the big data era, the massive, widespread collection, storage, processing, and circulation of data containing personal information creates enormous tension between the personal interests embedded in personal information protection and the massive economic interests, public safety, and public health contained

in information free flow and application [3]. This contradiction and conflict of interests will be further intensified in the big data era. How should the benefits generated by big data be distributed? How should we regulate and prevent the hidden dangers of personal information abuse brought by big data? How should we set the balance point when facing conflicts of interest?

Algorithmic Discrimination and the "Black Box" Problem

Scholars and governments are increasingly aware of the widespread existence of algorithmic discrimination and believe that these new forms of discrimination can trigger a series of social, ethical, and legal problems. While algorithmic automated decision-making brings convenience to people, it may also discriminate against certain groups due to the opacity of its decision-making process and information asymmetry. In practice, algorithmic discrimination mainly manifests in three basic forms: proxy discrimination by prejudice, discrimination by feature selection, and big data-enabled price discrimination against existing customers. Solon Barocas and Andrew D. Selbst argue that algorithmic automated decision-making may bring discriminatory adverse results to certain groups. In 2014, the White House released a report titled *Big Data: Seize Opportunities, Preserve Values* (Podesta et al., 2014). The report argued that influenced by the specificity of data sources and the subjective intentions of algorithm designers, algorithmic automated decision-making often creates hidden biases against applicants' employment, education, or credit. Such results can be self-reinforcing, systematically reducing individuals' access to credit, employment, and education, worsening their situation and placing them at a disadvantage in future applications.

Conclusion

We can see that big data is beginning to demonstrate its influence in legal services and legal research. This article does not advocate a kind of data fetishism. What big data brings will not be completely revolutionary change, but rather providing more refined intellectual support for decision-making, making decision-making and legal enforcement more efficient and targeted.

A more open question is whether the future of law will be replaced by endless data and algorithms? More than a decade ago, someone asserted that "code is law" in cyberspace. In the big data era, data and the algorithms that shape data value have become a new network architecture. While seemingly objectively analyzing massive amounts of data, this architecture also imposes this descriptive, ergodic fact as a normative rule upon everyone. What remains hidden behind are still the plans of different organizations and forces attempting to mine data value, especially when big data is widely applied in private transactions. Furthermore, this network architecture can be far more refined than existing law, thereby surpassing representative legislatures to propose new rules without any constraints. When algorithms become so complex that they

cannot be intuitively understood by humans, forming a "black box," this becomes dangerous. Therefore, the predictive use of big data analysis should not only fail to replace existing law but must also be constrained and supervised by democratic mechanisms—that is, controlled by democratic legislation. The key difference between law and technical architecture is that law reflects mainstream social values and is a product of compromise among diverse values and interests, whereas data mining serves only singular political or commercial interests. In the process of applying big data to legal practice and research, we should be mindful of this difference, thereby enabling big data to serve public interests and better align with legal practice.

References

[1] Zuo Weimin. "Toward Big Data Legal Research." *Legal Studies Research* 40, no. 04 (2018): 139-150.

[2] Zheng Ge. "The Law of Algorithms and the Algorithms of Law." *China Legal Review*, no. 02 (2018): 66-85.

[3] Mei Xiaying. "The Legal Attributes of Data and Its Position in Civil Law." *Social Sciences in China*, no. 09 (2016): 164-183+209.

[4] Liang Jiye, Feng Chenjiao, Song Peng. "A Survey of Big Data Correlation Analysis." *Chinese Journal of Computers* 39, no. 01 (2016): 1-18.

[5] Zhang Jiyu. "The Main Challenges and Opportunities Facing China's Judiciary in the Big Data Era—Also on the Demands of the Judiciary in the Big Data Era for Legal Research and Talent Cultivation." *Law and Social Development* 22, no. 06 (2016): 52-61.

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