Proceedings of ICILLP 2025 Symposium: Property Law and Blockchain Applications in International Law and Legal Policy
DOI: 10.54254/2753-7048/2025.LD29423

Risks and Mitigation Strategies for Generative Artificial Intelligence in Legal Document Drafting

Ximeng Zhou

School of Criminal Justice, Zhongnan University of Economics and Law, Wuhan, China
zxm18107119079@qq.com

Abstract. The breakthrough advancements of generative artificial intelligence (GAI) have catalysed a new wave of intelligent technology-driven revolution. Within the legal domain, GAI possesses extensive application potential in legal document drafting due to its robust natural language comprehension capabilities, precise data analysis skills, and highly efficient text generation. Consequently, inherent risks such as data security and algorithmic bias stemming from technical flaws in data and algorithms are increasingly evident. These risks further give rise to derivative concerns, including inadequate content accuracy, weak document explainability, difficulties in responsibility attribution, and deviations in ethical values. However, existing legal frameworks face application challenges in addressing issues surrounding GAI's use in public security legal document drafting, exhibiting supply-side inadequacies in data access, algorithmic review, and accountability mechanisms that struggle to keep pace with technological iteration. Only by coordinating three pathways—technological optimisation, institutional refinement, and human-machine collaborative governance—can the challenges of risk regulation in this domain be resolved. This approach unifies data security, algorithmic transparency, and controllable outcomes, thereby promoting the healthy and deep integration of generative AI with public security law enforcement operations.

Keywords: Generative Artificial Intelligence, Legal Documents, Application Risks, Regulatory Pathways, Smart Justice

1. Introduction

Legal documents are the statutory carrier for public security organs to carry out law enforcement activities. The quality of legal documents directly affects the following links of administrative penalty adjudication, fact finding and determination, evidence confirmation and legal application. The application of generative artificial intelligence overcomes the shortcomings of traditional manual drafting such as low efficiency, time-consuming and logical defects. It has certain significance in the production of public security legal documents and plays an important role in solving the problem of police burden and promoting intelligent policing. At present, academia has carried out in-depth research on the judicial application of artificial intelligence from the following perspectives: some scholars focus on building the macro risk regulation system; some scholars focus on specific problems in the judicial adjudication scenario, such as case fabrication and liability

© 2025 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: 10.54254/2753-7048/2025.LD29423

attribution, and put forward corresponding preventive measures; some scholars begin to explore the enabling mechanism and risk challenges of generative artificial intelligence (such as ChatGPT) in grassroots public security work. In general, existing research has made certain breakthroughs in the integration of generative artificial intelligence and the field of law, but it still has the following shortcomings: first, most of the current research focuses on the application of artificial intelligence in judicial adoptions or broad police work, while there is less special research on the scenario of drafting public security legal documents; Second, although some studies have mentioned the risk of artificial intelligence in police use, they cannot map the whole risk chain of artificial intelligence model training, input reasoning and output generation, considering the characteristics of public security legal documents. Third, the existing artificial intelligence risk mitigation approaches lack a governance system integrating legal guidance, technical regulation and dynamic supervision. They also ignore the special requirements of public security data protection and explainability of documents.

Given this, this paper explores how GAI can be applied in practice to help draft public security legal documents. By referencing advanced international application experiences and combining domestic examples such as Tuotian Public Security Large Model, we analyze application risks in different phases systematically, and then explore a collaborative governance road map adapted to public security law enforcement contexts, which combines technological and institutional levels. We aim to bridge the research gap at this level and offer theoretical references and practical guidance for the safe and compliant application of GAI technology in the policing field.

2. Technical advantages of generative AI in policing

In an increasingly rapid developing society, public security organs are faced with more complex criminal subjects and increasingly heavy demands for data processing. The core advantages of large-scale GAI models in policing are reflected in the following aspects: Firstly, strong language understanding and logical reasoning ability [1]. Large GAI models used in assisting the drafting of public security legal documents can accurately understand specialized terms and relevant statutory provisions. When dealing with complex case materials, they can quickly sort out the textual information and chain up the logic of the case quickly, and then deeply embed the applicable statutory provisions in the drafting of legal applications, making the content authoritative in terms of law and capable of persuading the judge in the application of law. Secondly, its ability to process multi-modal data. In addition to textual information, the GAI model also draws on body-cam footage, crime scene investigation photos and videos, interrogation audio recordings, etc. This multi-dimensional data integration method makes it possible to reconstruct the entire criminal process, and thereby improves the completeness and objectivity of the factual statements in the legal document. Thirdly, its ability to integrate and transfer knowledge from multiple fields. GAI integrated knowledge from multiple fields, legislative theory, criminological theory, criminal investigation, public security management, precedents, etc. This improves the accuracy and depth of the content of the facts stated, causal analysis and application of law in the document, and provides opportunities to break through the traditional investigation thinking in new criminal cases. Fourthly, continuous deep learning ability [2]. Based on the dynamically updated resources, including legal databases, case repositories and industry research, GAI continuously improves its algorithmic models and knowledge base to adapt to the changing criminal landscape and investigative needs, and ensures the timeliness of the technological application.

By leveraging the synergic effect of these advantages, large-scale GAI models can build a multi-dimensional case handling system, and the application of this system is an important technological

DOI: 10.54254/2753-7048/2025.LD29423

path to improve law enforcement efficiency and standardization.

3. Application scenarios of generative AI in police legal document drafting

In public security operations, the application of GAI in document drafting mainly covers two major scenarios: administrative law enforcement and criminal justice. Specifically, GAI assists in drafting both public security administrative documents and public security criminal documents. In administrative law enforcement, GAI is chiefly employed for the automated generation of legal documents such as administrative penalty decisions, administrative coercive measures documents, and case investigation reports. For high-frequency, low-complexity cases like traffic violations and market operation infringements, it generates standardised documents, markedly improving enforcement and penalty efficiency. Given the stringent requirements and considerable discretionary scope in administrative law enforcement, GAI applications must fully consider legality and fairness, striking a balance between efficiency and equity [3]. In criminal justice, GAI applications focus on critical stages such as case filing decisions, prosecution opinions, arrest warrants, and judicial rulings, exemplified by the Tuotian Public Security Large Model based on DeepSeek distillation technology. Potential risks in application—such as algorithmic black boxes generating false evidence or erroneous evidence assessments that undermine judicial credibility, data privacy breaches, and ambiguous accountability—must be prudently mitigated to safeguard the rigour and authority of criminal judicial documents [4].

In April 2024, Axon Enterprises' AI tool Draft One gained popularity among police officers for transcribing body-cam audio and automatically generating police reports during law enforcement operations. When US Police Sergeant Matt Gilmore and his police dog Gunner searched for suspects, the body-cam microphone captured their dialogue and the dog's barking. By extracting information from the audio, the tool generated a crime scene report in just 8 seconds – a task previously requiring 30-45 minutes. This not only resolves the tedious data collection, time-consuming on-site recollection, and labour-intensive documentation challenges of traditional methods, but also produces draft reports with precise, comprehensive details that require minimal revision after officer review, achieving exceptional efficiency. However, several significant risks warrant attention: Firstly, the issue of hallucinatory fabrication. When police reports serve as crucial evidence, their accuracy must be guaranteed. Yet the large language models underpinning AI are prone to fabricating information through hallucination, generating false content that may elude police detection and potentially exhibit biases such as racial prejudice [5]. Secondly, the issue of police retelling. As Draft One technology becomes increasingly widespread and officers grow more proficient in its use, their retelling to the machine may shift from detailed accounts highlighting key report points to rapid, verbose descriptions. This could lead to discrepancies between camera footage and actual events, potentially compromising the fair handling and adjudication of cases. Thirdly, the question of applicability. Police work possesses unique characteristics that brook no error. Draft One was initially designed for minor offence reporting. Its suitability for complex serious criminal cases, its application across all report types, and the deployment of visual recognition technology require further investigation.

4. Application risks of generative AI in assisting the drafting of police legal documents

The application of generative artificial intelligence to assist in drafting public security legal documents constitutes a comprehensive technological process encompassing the entire workflow from "data input to model computation to result generation".Application risks gradually spread and

DOI: 10.54254/2753-7048/2025.LD29423

materialize in three phases: model training, input inference and result generation. Risks inherent in these three phases have scenario-specific characteristics as well as cross-phase interdependencies [6]. Model training is the most fundamental phase for building GAI capabilities. Its inherent data and algorithm risks exist in three phases: model training, input inference and result generation. The input inference phase, which features human-machine interaction and information processing, will also experience recognition and understanding deviations. The result generation phase relates directly to the final quality and legality of documents. It will face challenges in terms of accuracy, explainability and responsibility for results. Accordingly, a systematic breakdown of these risks is presented below:

4.1. Model training phase

4.1.1. Data quality risk

The quality of training data has a decisive impact on model performance. In actual applications, the quality of training data varies greatly. Poor fitting of parameters in the model may occur, leading to factual errors in the generated results [7]. Lack of data representativeness may result in insufficient coverage of low-frequency case data, leading to deviation or distortion in non-typical case documents. If the training data used cover gender or regional discrimination in comparison with historical penalty records and sentences, content with systemic discrimination may be generated, leading to fairness risks when used.

4.1.2. Data security risks

Public security legal document training data involves a large amount of core data and sensitive information, including citizens' privacy and case confidentiality. Infringement and security incidents may occur due to leaks or misuse of data, and strict security protection should be provided for collection, storage and transmission process. If training data used includes unmasked privacy information, it may be misused by malicious attackers, leading to social risks. Data may also face multiple threats, including external hacking, internal police data misuse and cross-border transmission tampering [8].

4.1.3. Algorithmic bias risks

The black-box nature of algorithms inevitably gives rise to bias issues.The values of developers will be embedded, and the annotations of information will affect the model's interpretation of legal concepts, leading to subjectivity. If the training data used includes implicit discrimination against certain groups, the results generated by the system will inevitably be biased. For example, there may be descriptions with prejudice against certain professions or regions, leading to lack of judicial fairness [9]. In addition, the associative logic of parameters in GAIs deep learning models is unclear, and it is difficult to trace the cause of erroneous decisions [10].

4.2. Data entry and inference stage

4.2.1. Risk of information recognition errors

Whether information is manually entered by investigating officers or automatically recognised by intelligent systems, the input process inherently carries accuracy risks. Multimodal data recognition

DOI: 10.54254/2753-7048/2025.LD29423

is highly challenging and technically demanding. Issues such as illegible handwriting in written records, dialects and specialised terminology in speech-to-text conversion, and errors in surveillance footage extraction all impact case classification and proper handling. Furthermore, transcription must ensure information completeness; omitting critical details or including non-essential information undermines document standardisation.

4.2.2. Risk of information interpretation bias

Large language models (LLMs) are limited in their understanding of natural language. When faced with complex context or specialized terms, they are not able to distinguish the logical relationship between pieces of case information and make wrong inferences. As for some abstract legal concepts, LLMs cannot compare with investigating police officers in interpreting and judging according to facts in a certain degree. Therefore, when generating legal provision citations, content possibly contradictory to reality may be output. In addition, administrative laws and regulations are constantly changing, and models need to be updated in a timely manner with legal developments.

4.3. Result generation phase

4.3.1. Content accuracy risk

Public security legal documents demand strict adherence to content accuracy, logical coherence, and formal standardisation. Content generated by GAI may contain fabricated elements or logical contradictions. Ambiguous factual statements or erroneous legal reasoning will undermine the quality and persuasiveness of the document.

4.3.2. Document explainability risk

During automated generation, models may exhibit hallucinations due to insufficient training data, fabricating non-existent legal provisions or precedents. The complexity of algorithms renders model reasoning processes inherently less interpretable [11]. This introduces potential risks in document review, error correction, and utilisation. Failure to promptly identify and rectify such errors may directly erode the authority of legal documents and diminish judicial credibility in the eyes of the public.

4.3.3. Liability attribution risk

Defining responsibility in human-machine collaboration poses a significant challenge. When errors in documents cause adverse social impacts, developers may claim "merely providing technical tools" while users assert "reliance on model outputs," making fault allocation difficult. Furthermore, the issue of copyright ownership for algorithmically generated legal content remains unresolved and requires urgent regulatory clarification.

DOI: 10.54254/2753-7048/2025.LD29423

5. Mitigation pathways

5.1. Technical level: fortifying the foundations of intelligent applications

5.1.1. Optimising algorithmic models to mitigate algorithmic risks

To address algorithmic black-box risks, establish an explainable algorithmic framework by combining rule-based expert systems and deep learning models to improve the transparency and visualisation of GAI's reasoning logic [12]. Train models using adversarial training methods with Generative Adversarial Networks (GANs) to generate case scenarios for training models to discriminate against biased data. Quickly adjust model parameters to improve output and ensure the reasonableness of generated results. Embed core policing values—such as judicial fairness, procedural justice, the protection of human rights, and placing the public's interests first—into generative AI algorithmic models so that GAI can generate documents that conform to societal values based on the factual context of each case and ensure that value prioritisation is grounded in fundamental judicial principles [13]. Finally, enhance models' accuracy in inferring complex legal concepts to avoid fabrication based on hallucination.

5.1.2. Strengthening data management to ensure data security

Regarding data security, the public security system database should be comprehensively enhanced. Establish a rigorous, tiered data governance platform that classifies data into three levels based on sensitivity: core data, critical data, and general data. Data involving state secrets or national security should be subject to careful consideration before granting access to GAI. Differential encryption techniques should be applied according to data classification to prevent leakage and misuse throughout data collection, storage, and usage. During document generation, both the legitimacy of training data sources involving personal privacy and commercial secrets must be ensured, and dynamic data anonymisation should be implemented to mitigate risks such as algorithmic bias and infringement [14]. Regarding data quality, ensure comprehensiveness while screening and pre-processing collected training data. Update data promptly in response to legal and regulatory changes to adapt to evolving requirements in public security legal document drafting. Employ blockchain evidence storage technology to achieve full traceability across data sourcing, annotation, training, and archiving. Conduct regular data quality audits to sustain long-term data integrity and accuracy.

5.2. Legal and institutional framework: establishing regulatory operating guidelines

By refining laws and regulations and establishing oversight systems, accelerate the formulation and implementation of management regulations governing the application of generative AI in policing. Clarify the legal status, validity, and scope of application of GAI in drafting legal documents to ensure its operation strictly within the boundaries of the law. Define the boundaries of responsibility for both technology developers and users [15]. Algorithm developers must ensure model transparency and security to prevent technical vulnerabilities. Public security departments bear the responsibility to review the legality and accuracy of documents automatically generated by GAI, promptly correcting any errors.

At the regulatory level: Firstly, build a comprehensive supervision system to supervise the application of AI in police legal document drafting in the whole process. Before generation, strictly check whether the algorithm is lawful and reasonable; if there are designing discriminatory phenomena, they should promptly require correction. In the process of generation, users must do the

DOI: 10.54254/2753-7048/2025.LD29423

real-time supervision and correct to show at any time the procedural error. After generation, the police staff should do the final verification. The police should review the content of the document and correct the document to make the document content fully comply with objective facts of the case. Secondly, develop an AI policing application registration system. Specialized technical inspection institutions shall conduct compliance testing on all intelligent systems that have been deployed in use to avoid the phenomenon of algorithmic bias and privacy violation caused by the application of algorithms [16]. Thirdly, convene technical experts, legal experts and police staff to formulate industry standards for GAI applications in the public security system together. The standard shall be updated in time according to the changes in the crime trend, and shall make unified technical rules and ethical guidelines.

5.3. Human-machine collaboration: achieving efficient and secure cooperation

Clarify that Police officers play the main role in drafting public security legal documents, while generative AI serves an auxiliary role. As an auxiliary tool, generative AI should focus on basic information retrieval and drafting the first version. It should leverage advantages in handling large quantities of data and integrating information from multiple sources to improve efficiency in generating documents. Police officers are the legally responsible persons for drafting legal documents. They should do the overall review and verification of the document generated by GAI. It will be factually correct and strictly implement the law in applying laws to ensure the legal power and social justice. Those who are responsible for supervision during the review shall be strictly held accountable for the problems caused by supervision during the review. An intelligent correction feedback system has been constructed and applied. After substantive review and amendment by legally qualified officers, the officer's amendment will be automatically fed back to the training end of the generation model to adjust the model parameters to achieve better optimization.

Simultaneously, follow the principle of "technology should help, not replace human work". Introduce specialized training for police officers in using generative AI as a work assistant, enhancing the standardized use of intelligent tools by police and the ability to authenticate generated content. Focus on digital literacy training based on the operation capability of AI tools and identification of algorithmic risks. Continue to collect user feedback from police personnel on the use of intelligent cooperation system, continuously improve the human-machine interface design and application flow of substantive review, further improve system performance and user experience, and continuously meet the practical needs of police in drafting legal documents.

6. Conclusion

The deep integration of generative artificial intelligence with legal drafting represents a landmark manifestation of technology empowering standardised law enforcement and judicial modernisation. This paper systematically discusses the application advantages and scenarios of GAI in drafting public security legal documents, and further explores the various risks that may occur in training, inference and generation processes of GAI. Based on this, this paper proposes a closed-loop risk mitigation path combining technological optimisation, institutional construction and human-machine collaboration. Future efforts should focus on deepening exploration in the following areas: Firstly, advancing the implementation of explainable AI technologies within legal contexts to enhance algorithmic decision-making transparency and auditability. Secondly, refining the legal regulatory framework for generative AI services by establishing end-to-end oversight mechanisms covering data access, algorithmic registration, and output verification. Thirdly, strengthening human-centric

DOI: 10.54254/2753-7048/2025.LD29423

collaborative models through training to elevate officers' digital literacy and risk identification capabilities, while clarifying their primary responsibilities in document generation processes. This will enable the evolution of AI-assisted legal document drafting from mere usability towards trustworthiness, reliability, and accountability, thereby providing robust support for the development of intelligent justice systems.

References

[1] Wang Lusheng. From Evolution to Revolution: The Legal Profession Under the Influence of ChatGPT-like Technologies [J]. Eastern Chinese Law Review, 2023, (04): 56-67.
[2] ALEX T, MILES B, JACK C, et al.Understanding the capabilities, limitations and societal impact of large language models [J].ar Xiv, 2021(2).
[3] Wei Qiong, Xu Junhui. Application and Legal Regulation of Artificial Intelligence in Administrative Penalties [J]. Times Law Review, 2021, 19(01): 12-22.
[4] Wei Bin. Challenges and Pathways for Integrating Judicial Artificial Intelligence into Judicial Reform [J]. Modern Jurisprudence, 2021, 43(03): 3-23.
[5] Yao Li. Risks and Implications of ChatGPT Fabricating Legal Precedents: An Analysis of the 2023 US Case "Mata v. Avianca" [A]. Emerging Rights Anthology 2024, Vol. 2 — Research Collection on Emerging Rights in the Context of Artificial Intelligence [C]. Shanghai Law Society, 2024: 11.
[6] Liu Yanhong. Three Major Security Risks of Generative Artificial Intelligence and Legal Regulation: The Case of ChatGPT [J]. Eastern Chinese Law Review, 2023, (04): 29-43.
[7] Zhang Xiaoyan. A Review of Generative AI Data Research: Risks, Challenges and Governance [J]. Library and Information Work, 2025, 69(09): 136-148.
[8] Weng Sunzhe. Risk Identification and Regulatory Pathways for Embedding Artificial Intelligence Algorithms in Public Security Law Enforcement [J]. Journal of Political Science and Law, 2023, 40(03): 17-25.
[9] Xu, E. P. Risks and Solutions for Generative Artificial Intelligence Applications in the Judicial System [J]. China-Arab Science and Technology Forum (Chinese-English Edition), 2024, (01): 149-155.
[10] Marcelo Pasetti, James William Santos, Nicholas Kluge Corrêa, Nythamar de Oliveira, Camila Palhares Barbosa. Technical, legal, and ethical challenges of generative artificial intelligence: an analysis of the governance of training data and copyrights [J]. Discover Artificial Intelligence, 2025, 5(1): 193-193.
[11] Liu Yanhong. Research on the Explainability of Artificial Intelligence and the Issue of AI Legal Liability [J]. Law and Social Development, 2022, 28(01): 78-91.
[12] Zheng Xi. The Application of Generative Artificial Intelligence in the Judiciary: Prospects, Risks and Regulation [J]. Chinese Journal of Applied Law, 2023, (04): 81-93.
[13] Xie, Dengke. Challenges and Transformations in AI-Driven Digital Prosecution [J]. Journal of China University of Political Science and Law, 2023, (06): 231-241.
[14] Shang Jiangang. Research on Meta-Rules for Generative AI Risk Governance [J]. Eastern Chinese Law Review, 2023, (03): 4-17.
[15] Jiang Ye. The Discipline of Algorithms and the Algorithms of Discipline: Legal Regulation of Algorithms in the AI Era [J]. Hebei Law Review, 2018, 36(12): 142-153.
[16] Guo Chun-zhen. Coherent Legal Governance of Generative AI: Taking Generative Pre-trained Models (GPT) as an Example [J]. Modern Jurisprudence, 2023, 45(03): 88-107.