



Centro de Estudios de Justicia de las Américas
Justice Studies Center of the Americas

Artificial Intelligence in Public Prosecution: Reflections and Guidelines for the Americas

www.cejamericas.org



© 2026 Justice Studies Center of the Americas (JSCA/CEJA)

San Pío X 2460, Office 608, Providencia, Santiago, Chile

Phone: +(56 9) 5611 2162

www.cejamericas.org

Title: *Artificial Intelligence in Public Prosecution: Reflections and Guidelines for the Americas*

First Edition, 2026

Copyright and ISBN: Pending.

Author

Pablo Pizarro Zúñiga

Project Team

Nataly Ponce Chauca

Rodrigo Nakada Castro

Kevin Gutierrez Parra

This publication is distributed free of charge in both print and digital formats.

The reproduction of this publication, in whole or in part, is authorized provided that the source is properly acknowledged

Introduction	3
I. Principles for AI governance and implementation in Public Prosecution	5
Principles	5
II Recommendations for Establishing Artificial Intelligence Governance in Public Prosecution	10
III. Risk stratification for institutional control and the protection of fundamental rights	14
Prohibited uses	15
High risk	15
Low risk	16
IV. Artificial Intelligence for the Management of Public Prosecution	17
Tasks related to case intake and classification	17
Tools for assisting victims, users, and litigants	20
Transcription, Digitization, and Non-Case-Specific Applications	21
AI for Institutional Management and Oversight	22
V. Decision Support in Criminal Investigation and Prosecution	25
Improvements in Complex Criminal Investigations	25
Malicious Uses of Artificial Intelligence in Criminal Activity	26
Litigation and Judicial Activity Analysis	27
Risk Assessment of Recidivism and Criminal Offending	30
Strengthening Compliance with Institutional Guidelines	34
Conclusions	36
References	38

Introduction

Public Prosecution Offices, responsible for criminal investigation and prosecution, face an operational paradigm shift driven by the rapid advancement of digital technologies and the emergence of Artificial Intelligence (AI). These institutions are integrating technological solutions not merely as administrative aids, but as strategic assets to navigate the increasing complexity of crime, emerging typologies, and high caseloads. The goal is to leverage AI to optimize case management, accelerate response times, and significantly enhance the effectiveness of criminal prosecution.

Current deployments encompass a broad spectrum of critical functions, ranging from digital interfaces that streamline the protection, guidance, and intake of victims and witnesses, to advanced criminal analysis systems capable of mapping complex criminal phenomena. Solutions now include automated recidivism risk assessments, high-volume digital evidence processing, and intelligent document management. Furthermore, agencies are adopting mechanisms for complaint pre-classification, strategic data utilization, and applications that assist in assessing legal viability and drafting court documents.

While the operational benefits are evident, integrating AI into criminal investigations presents significant risks to procedural safeguards and fundamental rights. Without rigorous oversight, the use of AI by prosecutors and supervised investigative bodies threatens to reproduce or amplify systemic biases against specific population segments, compromise the objectivity of investigations, and undermine the presumption of innocence. Such failures directly impact citizens' rights and the overall effectiveness of prosecution. Moreover, there is a tangible risk that automation could displace or minimize essential human judgment, create evidentiary asymmetries for the defense, or perpetuate obsolete and discriminatory decision-making criteria based on nationality, gender, race, or socioeconomic status.

These internal challenges are compounded by an escalating external threat: the malicious use of generative AI to alter reality with high sophistication. This capability undermines the reliability of

traditional forms of evidence and jeopardizes the ability to prove criminal liability, constituting an unprecedented institutional and systemic challenge to the administration of justice.

This report organizes its analysis and proposals around a robust institutional framework designed to harness AI benefits without compromising legality, objectivity, or due process guarantees. The document is structured as follows: first, it establishes guiding principles for AI governance and implementation in Public Prosecution Offices; second, it details the necessary governance architecture, identifying minimum institutional elements to ensure control, traceability, and accountability, and finally, it examines specific applications for administrative management, investigative effectiveness, and decision-making support.

I

Principles for AI governance and implementation in Public Prosecution

There is still a lack of a comprehensive regulatory framework governing the use of AI in Prosecution Offices and District Attorney's Offices. Similar to the judicial sphere, there is an absence of institutional governance structures, technical guidelines, and guiding principles that clearly define the scope, limits, and conditions of using these technologies for tasks that may affect fundamental rights. This allows for a wide margin of discretion in implementation and amplifies the associated risks, especially in contexts of operational pressure, human resource shortages, or systems with weak internal oversight.

Adopting explicit principles is essential to ensuring the stable incorporation of these automated systems and regulating their governance and implementation. Below is a non-exhaustive catalog of principles relevant to the implementation of AI in public prosecutor's offices, adapted from the most frequently reiterated principles for the incorporation of AI in the judicial sphere (JSCA 2025, [UNESCO 2025](#)).

Principles

Human Control and Non-Replacement of Human Rationality

AI tools may be used to support administrative management and tasks related to investigations and criminal prosecutions, provided that appropriate safeguards are adopted in view of the risks they pose. AI tools may never fully automate or replace decisions explicitly assigned to prosecutors by law, including formalizing, charging, requesting precautionary measures, negotiating alternative outcomes, and any other decision that could affect the fundamental rights of those involved. Likewise, humans must always review, evaluate, and critically correct, when appropriate, the results and recommendations generated by these tools.

Primacy of fundamental rights and due process

The use of AI in the criminal system must prioritize safeguarding the fundamental rights and guarantees of defendants, victims, and other involved parties. Consequently, the design, implementation, and use of AI may not undermine the presumption of innocence or infringe upon the right to defense, equality before the law, effective judicial protection, or the rights to participation, protection, and non-revictimization of victims.

When AI is used to make substantive decisions in criminal proceedings, the due process rights of defendants, victims, and other involved parties must be protected. In particular, the results and rationale must be understandable and reviewable; sufficient traceability must be ensured to verify its functioning, and effective mechanisms must be in place to challenge its use and adverse effects.

Objectivity and respect for the presumption of innocence

Prosecutorial bodies are bound by an institutional obligation to act objectively. Consequently, the AI tools employed during the investigative stage must be configured and utilized in accordance with this principle. In particular, the court must identify and make available relevant evidence to prove punishable participation and to support hypotheses of the defendant's innocence, giving equal weight to both components.

Prevention of bias, non-discrimination, and mandatory prior testing

AI systems that affect case processing must be analyzed and tested before final implementation. Once operational, they must undergo periodic algorithmic impact assessments to prevent the reproduction or amplification of historical biases. Measures must be taken to prevent biases associated with gender, race, neighborhood, socioeconomic status, or membership in vulnerable groups so as to prevent selective prosecution, the disproportionate use of precautionary measures, and the unjustified denial of alternatives to trial.

Transparency and explainability

The use of AI must adhere to transparency standards that ensure democratic control and safeguard guarantees. Individuals subject to automated decisions or those made with AI tools must be informed promptly and clearly of the existence and role of these tools in the process. This information must be compatible with due process, investigative confidentiality, and security. Likewise, institutions must publish algorithmic impact assessments, pilot test results, and key background information on the operation, objectives, limitations, and performance of the systems used in an accessible format that protects sensitive information. This allows internal and external bodies to examine and evaluate their adoption.

Auditability and traceability

All relevant uses of AI must be auditable and traceable. Consequently, they must leave a verifiable record allowing for the reconstruction of, at a minimum, the input data used, the model version employed, the generated result or recommendation, the performed human review, and the ultimately adopted decision. These requirements are essential for internal control, subsequent review, and determining institutional responsibilities, especially for decisions impacting cases.

Accountability and Responsibility

The use of AI requires a clear system for assigning accountability and responsibility. Consequently, the responsible parties for each phase of the system's life cycle must be defined, including design, deployment, maintenance, and updating. Additionally, the parties responsible for authorizing and validating uses must be defined, as well as the appropriate measures in the event of failures or unforeseen effects, such as correction, suspension, deactivation, and, where appropriate, repair. This principle aims to prevent the diffusion of responsibility that is common in complex socio-technical systems and ensure effective institutional control.

Data Protection and Privacy

The use of AI must comply with current national and international legislation on personal data protection and privacy. Reinforced standards must be adopted for sensitive information. Special safeguards must be adopted regarding data on victims, witnesses, and whistleblowers to ensure confidentiality, minimize data, and restrict access. Similarly, institutions should use AI to protect the privacy of third parties when their data is part of an investigation. They should also establish safeguards to prevent prompts or inputs from exposing information outside institutional infrastructure, especially when using external tools or third-party services.

Training, education, and informed use

Implementing AI requires ongoing training tailored to different profiles and responsibilities to enable a critical understanding of limitations, errors, and biases. The goal is to avoid automation bias and ensure that human review is substantive, not merely formal.

Collaborative governance

The definition of uses, limits, and protocols for AI should be carried out through an interdisciplinary approach that incorporates legal, technical, and management knowledge. This approach should also include the participation of units with expertise in victimology, data protection, and security. This collaborative process improves design, reduces resistance, and strengthens effective oversight.

These principles cannot resolve the dilemmas posed by incorporating AI into investigative and prosecutorial functions. Nevertheless, they have operational value as they serve as a frame of reference for evaluating the appropriateness of particular implementation measures.

The next step is to define governance mechanisms that enforce these principles in practice. This includes defining how competencies are distributed, how risk is stratified, and the regulatory implications. It also includes other relevant elements of the institutional architecture necessary to introduce these innovations.

II

Recommendations for Establishing Artificial Intelligence Governance in Public Prosecution

AI provides concrete opportunities for prosecution offices to enhance efficiency, institutional coherence, and strategic capacity in criminal prosecutions. However, this potential can only be realized if it is accompanied by robust governance frameworks that ensure the ethical use of AI in accordance with current regulations, the democratic rule of law, and human rights.

Even when AI supports human reasoning without replacing it, it can significantly influence prosecutors' reasoning and actions because it prioritizes and provides information that shapes decisions determining an individual's freedom or imprisonment. Without adequate institutional mediation, there is a risk of consolidating undetected biases, straining professional autonomy, and shifting discretion to less visible areas, such as data selection, parameterization, or coding. This weakens accountability and effective control.

Formalization of institutional governance of AI

The adoption of AI in public ministries should be treated as an institutional policy decision, not a series of isolated technological initiatives. When the design, procurement, or use of tools is left to discretionary decisions, variability increases regarding what is permitted, the applicable standards, and the applicable controls. It then becomes more likely that substantial impact

technologies will operate without sufficient checks and balances. In contexts of high workloads, weak internal controls, and scarce resources, the incentive to implement "quick" solutions can override the assessment of legality, proportionality, and guarantees. This increases the risk of practices that are incompatible with due process and weakens institutional accountability (Vargas-Murillo et al 2024).

Formalizing governance means establishing rules in an official document that governs the entire life cycle of systems. This process places a central emphasis on the distribution of powers and responsibilities. It identifies who has the authority to promote, approve, deploy, and maintain a system, what minimum standards must be met, what controls are required before and after implementation, and who is accountable for failures or adverse impacts. Without this structure, transparency and human oversight tend to be merely declarative, not effective conditions for control (Coglianese & Lehr 2018; McGregor et al 2019).

Authority and roles

Clear roles and formal authority are required. Examples include governance committees and specialized leadership positions with the power to validate projects, demand evidence of performance, approve pilots, and authorize scaling up initiatives for final implementation. This distribution reduces discretion and brings consistency to the adoption of AI tools.

Mandatory protocols and life cycle

Governance must be translated into mandatory operating protocols that standardize processes such as selection, design, procurement, pre-evaluation, implementation, monitoring, auditing, and deactivation. Standardization is essential for ensuring institutional continuity and consistency between units, as well as the ability to effectively monitor performance and risks over time. Governance must cover the entire life cycle and include continuous evaluation, technical audits, clear documentation, and sustained control mechanisms.

Responsibility, human oversight, and traceability

An explicit accountability regime must be established to ensure meaningful human oversight, traceability, and documentation. This allows us to reconstruct what the system did, what was reviewed, by whom it was reviewed, and what decision was ultimately made. This is particularly important when AI influences sensitive decisions because control deficits can undermine safeguards and hinder accountability.

Risk stratification as the core of governance

A structural component of formalization is risk categorization. Since not all applications require the same level of scrutiny or safeguards, governance must incorporate a classification process that distinguishes between low-, medium-, and high-risk uses. There should also be a category for prohibited uses when the potential impact on rights or guarantees is not justifiable. This stratification allows each level to be linked to specific requirements, such as prior algorithmic impact assessments, external audits, and explainability conditions, among others.

Benefits of Formalizing Governance

Formalizing governance limits risks and allows benefits to be captured sustainably. With rules in place for adoption, inventories, prior evaluation, pilot testing, monitoring, and traceability, an institution can more consistently incorporate AI, avoid duplication, integrate solutions, learn from failures, and publicly uphold the legitimacy of its practices.

The technology provision model as a key decision

A key decision for AI governance is choosing the technology provision model. AI can be developed internally, outsourced to a company according to specific requirements, obtained through inter-institutional collaboration, or general-purpose commercial chatbots can be authorized (JSCA 2025).

The choice of technology provision model affects operational autonomy, traceability of systems, and most importantly, the degree of control that can be exercised over applications affecting critical prosecutorial management functions. When a prosecutor's office adopts technologies, whose operation cannot be audited or whose logic remains opaque, even to users, the office limits its ability to review, correct, or justify decisions influenced, guided, or recommended by the tools. Consequently, a significant part of the risk comes from the structural conditions under which the outputs are produced and operated, not just the outputs themselves.

For this reason, the technology provision model decision must consider the effective possibilities for auditing, access to relevant documentation and parameters, and the capacity for adjustment and correction.

III

Risk stratification for institutional control and the protection of fundamental rights

AI applications do not present a uniform risk, so an essential component of robust governance is risk stratification. This involves classifying uses according to their potential impact and, based on that classification, proportionally adjusting validation, oversight, and accountability requirements. The goal is not to stifle innovation, but to ensure that institutional deployment maximizes benefits without compromising safeguards, legitimacy, and control.

A functional stratification model that has been observed in various judicial rules, following the example of the risk-based model in the [European Union's 2024 AI Act](#), is the distinction between different types of use.

The decision on risk is primarily regulatory, i.e., in cases where there is a legal framework regulating uses or establishing parameters; the institutional decision must comply with that framework. When no such framework exists, it will be an institutional decision, in which governance mechanisms must interact to generate a cross-cutting, open, and participatory institutional reflection in order to adopt the stratification strategy.

Below is an example of risk stratification based on the functions typically performed by public prosecutors.

Prohibited uses

Incompatible applications, either because they are openly contrary to the principles set out above, for example, because they lack human control, because they replace prosecutors in substantive decisions, or because they involve a level of infringement of rights that is not proportional to the achievement of the desired objectives. Example: systems that automatically determine the risk of recidivism without human supervision.

High risk

AI systems that influence investigative decisions or the exercise of prosecutorial powers, such as supporting legal classification and charges, recommending or supporting requests for precautionary measures, guiding decisions to dismiss or prosecute, or prioritizing criminal cases and phenomena, are considered high risk. These uses include supervised prediction models that, based on historical data and case variables, generate scores or recommendations on criminal risk or the probability of certain procedural outcomes.

This also includes applications that have an impact on the rights of the parties, such as early decisions on the fate of cases (e.g., on admissibility, level of complexity, whether to dismiss the case for lack of evidence, dismissal for not compromising the public interest, statute of limitations, among others).

Similarly, systems that systematize information, summarize it, or provide procedural and legal forecasts that prosecutors can use to build their case theory or determine the most appropriate procedural strategy to apply also represent a high risk.

Similarly, automated systems that communicate relevant case information to victims of domestic violence or sexual crimes are considered high risk in cases where the accused shares a residence with the victim or is in such close proximity that there is a plausible risk of digital identity theft.

Low risk

This category includes administrative or internal support tools that have no direct impact on legal decisions or fundamental rights (e.g., scheduling, reporting, anonymization, data analysis, digitization, etc.). At this level, basic traceability requirements and periodic reviews are usually sufficient.

Ultimately, the greatest benefit of stratifying risks lies in the ability to organize the AI ecosystem using a proportionality criterion: the greater the potential impact, the greater the control requirements. This helps to protect rights and uphold institutional legitimacy, without sacrificing the operational and strategic benefits offered by AI.

IV

Artificial Intelligence for the Management of Public Prosecution

Tasks related to case intake and classification

Potential advantages

The automation of administrative support tasks, i.e., those that are routine and do not contribute substantive legal value, frees up critical institutional resources. This allows prosecutors and legal and administrative staff supporting prosecutions to reduce their workload associated with bureaucratic procedures and repetitive tasks of low legal value, so that they can devote most of their efforts to investigation and criminal prosecution¹. With regard to the pre-classification and early management of complaints, AI makes it possible to substantially optimize the control of case intake and processing. For example, through natural language processing, it is possible to automate the preliminary analysis of the facts described in police reports or criminal complaints. This allows AI systems to identify the type of crime, predict the level of complexity and viability of the case, and recommend preliminary proceedings according to instructions or standards

¹ Many of the potential advantages of incorporating AI into the administrative sphere of prosecutors' offices or public ministries are identical to those that can be observed in its incorporation into the judicial sphere. Therefore, for more details on the potential uses of AI tools in the management of justice services and specific activities, we recommend consulting the JSCA document "Artificial Intelligence in the Judiciary: Reflections and Guidelines for the Americas" (2025).

previously established by the prosecutor's office. It also allows for the implementation of automatic referral processes or suggests the dismissal of the case, either because the criminal action is time-barred, because the facts do not constitute a crime, or for any other reason provided by law.

The use of AI tools in support tasks that do not affect the processing of criminal cases is considered low risk or risk-free (such as scheduling, digitization, thematic classification, workload reports, etc.), and their incorporation and implementation is highly recommended, as they have the potential to bring high gains in efficiency with low risk of undermining fundamental rights.

Risks

However, when it comes to using automation processes to assist in “low-complexity” decision-making that impacts the outcome of prosecutions—such as dismissing a complaint due to the statute of limitations, proposing preliminary proceedings in less serious crimes, deciding to close a case due to lack of evidence, or determining that the facts contained in the complaint do not seriously compromise the public interest—the use of AI tools affects the decision-making process for actions that have legal consequences for the accused and the victims. Therefore, their design and adoption must take these potential impacts into account. Furthermore, the vast majority of these decisions are placed by law under the non-delegable responsibility of prosecutors, even though in practice many of them are made by administrative or support staff.

This raises the question of whether the use of these tools affects the legal validity of these decisions. In particular, it is worth asking whether decisions made wholly or partly by AI systems, in the absence of a specific legal basis enabling them and sufficient guarantees of transparency, traceability, and meaningful human control, meet the minimum standards of legality, motivation, and competence required of the prosecutorial function by general principles of law and current regulations. If the decision lacks real intervention by a prosecutor, or if it is based exclusively on unverifiable automated reasoning, it could suffer from procedural or substantive flaws that affect its validity and make it potentially contestable, either administratively or judicially, affecting

criminal prosecution, the right to defense, and, from a broader perspective, the legitimacy of the criminal process.

In a scenario of inorganic and accelerated adoption of AI, in which many prosecutors' offices in the region use these tools without institutional governance mechanisms or national regulatory frameworks for their responsible adoption, human supervision has been presented as a safeguard against these risks, either as a principle of the system or as an attribution of responsibility for the duty to review the outputs generated by AI tools.

This raises another question for public prosecutors: is nominal human oversight really a safeguard for fundamental rights? In the high-workload contexts in which prosecutors in the region operate, such oversight is limited or downright impractical. And even if there were supervision, its effectiveness would be diluted without the technical and organizational means to critically review the “reasoning” of the AI tool used.

Therefore, the development and application of AI at this stage of the criminal process, even for routine decisions, must be approached with an institutional focus that considers the risks involved in these uses and keeps in mind the principles that govern the work of prosecutors: investigative objectivity, legality, proportionality in their actions, public oversight, and the guarantee of fundamental rights. Without clear regulation and control mechanisms, even “easy” decisions can affect the democratic foundations of the criminal justice system.

Tools for assisting victims, users, and litigants

Potential advantages

The incorporation of virtual assistants and chatbots in prosecutors' offices and public ministries offers concrete opportunities to strengthen interaction with those who are part of the criminal justice system. For users in general, particularly victims and complainants, these digital channels

can expand the possibilities for contacting the prosecutor's office on non-business days and hours, overcome geographical barriers, and improve access to basic procedural information, or serve as a communication channel for bringing key or urgent information to the attention of the institution.

For legal professionals, these tools can significantly improve institutional response times, facilitate access to documentation, enable the coordination of remote interviews with prosecutors, and streamline the request or follow-up of proceedings. This not only optimizes the operational workload of institutions but also improves the experience of those who must interact regularly with the system in the exercise of defense, representation, or legal advisory functions.

Risks

However, when these solutions are not designed with a rights-based perspective, they can have adverse consequences for victims, witnesses, and complainants that must be anticipated and mitigated. One of the main risks relates to the privacy and protection of victims' personal data. In situations of gender-based violence, domestic violence, or in contexts of cohabitation with the aggressor, the use of automated systems without robust authentication and control mechanisms can facilitate unauthorized access to confidential information. The aggressor could impersonate the victim, obtain details of the status of the process, or intercept sensitive communications, which not only constitutes a threat to personal safety but can also hinder the investigation or give rise to new forms of intimidation or violence.

Another critical factor is the risk of digital exclusion. The implementation of technological tools as the only means of assistance can widen the gap in access to the justice system if the social, economic, or technological conditions of vulnerable groups are not taken into account, as they may face significant obstacles in interacting with these platforms.

Without in-person alternatives or institutional support, these solutions could become a new barrier to access, contrary to the principle of equality.

Therefore, despite being a priori low-risk, use in terms of the impact on fundamental rights, their implementation must necessarily adopt this perspective in order to prevent adverse effects.

Transcription, Digitization, and Non-Case-Specific Applications

Potential Advantages

In the domains of transcription, digitization, and inter-institutional communication, AI tools offer significant opportunities to modernize investigative procedures. These technologies enable the digitization of physical records, the automated transcription of witness statements and hearings, and the generation of standardized drafts for information requests and routine administrative resolutions. Since these functions do not directly influence the outcome of criminal cases, they serve as high-impact solutions for improving the quality, uniformity, velocity, and traceability of institutional actions.

The transition toward a fully digital documentary ecosystem—supported by AI—is fundamental to enhancing access to justice. This shift facilitates remote hearings, centralizes legal and administrative data, and improves record preservation, security, and system interoperability. Furthermore, automated processing for format conversion allows institutions to standardize document structures, correct formal errors, and enhance legibility.

A critical contribution of these applications lies in the protection of fundamental rights, particularly regarding privacy and data confidentiality within investigative files. AI-driven automated anonymization systems can identify and redact sensitive information concerning victims, defendants, or third parties before records are disclosed or shared between agencies.

Reflecting this technical nature, both the [European Union's AI Act](#) and the [guidelines of the National Council of Justice of Brazil](#) categorize these uses—such as anonymization, document standardization, and automated transcription—under regulatory frameworks with less stringent obligations.

AI for Institutional Management and Oversight

Potential advantages

The implementation of AI in management oversight represents a strategic opportunity for public prosecution offices to achieve more efficient resource administration. These applications focus on internal monitoring, supervision, and planning.

Primarily, AI facilitates the proactive monitoring of judicial and administrative deadlines by generating automated alerts for pending actions. Integrating case management systems with institutional calendars reduces the risk of procedural omissions. Additionally, AI models can detect delays in tasks assigned to police forces, technical experts, or internal units.

AI also enables a more equitable and efficient distribution of workloads. By analyzing data on active cases, estimated legal complexity, and team performance history, systems can generate dynamic assignments that prevent the overburdening of specific units or prosecutors. Furthermore, these technologies systematize key information for strategic decision-making

through the automated generation of reports and dashboards. While categorized as low-risk, these benefits must be weighed against potential organizational resistance.

Risks

Research indicates that incorporating AI for tracking and control may yield unintended consequences for personnel, such as increased administrative burdens, the perception of permanent surveillance, or a diminished sense of professional autonomy. [Bayne & Christin \(2020\)](#) observed that algorithmic systems in criminal justice, if implemented without adequate institutional mediation, can intensify hierarchical control, reduce professional discretion, and increase work-related stress.

To mitigate these adverse effects, AI implementation should involve participatory processes, comprehensive training, and clear regulations regarding the scope and purpose of the tools. These systems must be framed as technical support for human decision-making rather than punitive evaluation mechanisms. A responsible and transparent deployment is essential to ensuring institutional efficacy and acceptance.

V

Decision Support in Criminal Investigation and Prosecution

Improvements in Complex Criminal Investigations

The expansion of AI-based technologies provides an opportunity to augment the capacity of prosecutorial agencies to manage the increasing complexity and sustained volume of criminal activity in the region. Its ability to process massive datasets, combined with its efficacy in detecting latent patterns and generating strategic recommendations for prosecutors, achieves results that would otherwise require prohibitive amounts of time and human resources. One of the primary weaknesses of traditional prosecution is information fragmentation. AI facilitates the identification of recurring criminal modalities across vast quantities of case files, fostering criminal analysis and inter-jurisdictional investigative focuses. This transition allows for a shift from case-by-case investigations toward the strategic prosecution of criminal markets, resulting in significant efficiency and quality gains across various criminal domains.

In the field of financial investigations and money laundering, AI enables the tracking of capital through thousands of banking, corporate, and tax transactions. Machine learning and anomaly detection algorithms can identify structuring schemes, complex financial triangulations, or hidden links to shell companies undetectable to human analysts, significantly reducing false

positives in the analysis of suspicious activity reports ([FATF 2021](#)). Similarly, in the fight against child sexual exploitation and the storage of pornographic material, perceptual hashing tools allow for the automated classification and analysis of millions of image and video files, identifying illicit material and accelerating the identification of victims and perpetrators. Additionally, automated biometric image analysis facilitates matching against victim databases, aiding in the discovery of sexual exploitation networks ([EUROPOL 2023](#)).

Regarding phenomena such as human trafficking and migrant smuggling, natural language processing enables the correlation of thousands of recruitment advertisements published by criminals, automatically identifying relevant data such as phone numbers, semantic patterns in offers, and geolocation data ([OSCE 2020](#)). In forensic genetics, AI offers a critical solution for interpreting complex DNA mixtures; through machine learning algorithms, it is possible to separate and identify individual genetic profiles within contaminated samples or those with multiple contributors ([NIJ 2018](#)). Furthermore, in facial recognition, AI acts as high-speed support capable of identifying subjects under adverse conditions—such as difficult angles, low lighting, or covered faces—as well as reading license plates in blurred videos and deciphering characters illegible to the human eye (NIJ, 2018). The aggregation of these technologies provides public prosecution offices with new tools to counter the operational sophistication of organized crime, as these AI applications detect hidden patterns and complex correlations that allow for more effective mapping of the hierarchical structures of illicit organizations ([EUROPOL 2024](#)).

Malicious Uses of Artificial Intelligence in Criminal Activity

The phenomenon of AI represents both a strategic opportunity for strengthening judicial management and a driver of sophisticated criminal activity that poses critical challenges for public prosecution offices. The integration of these technologies into criminal acts compels

prosecutors to rethink their capacities for detection, fraud verification, and the substantiation of punishable facts. This reality demands an adaptive institutional response where the same technology that sophisticates crime serves as the foundation for bolstering criminal investigations (Ligeti et al 2025). Within this logic, the massification and refinement of cybercrimes—such as electronic fraud, phishing, and personal data theft—have reached a level of perfection that complicates user detection and significantly increases the probability of victimization. AI's ability to generate high volumes of attacks with extreme personalization multiplies the risk of deception on a global scale.

Furthermore, the use of generative AI tools to create deepfakes challenges the reliability of traditional forms of evidence. This dynamic of sophisticated forgery casts doubt on the veracity of documents, credit instruments, and overall, any medium of proof whose authenticity was previously unquestioned. This represents an unprecedented obstacle for prosecutors when establishing criminal participation and constructing a solid theory of the case before the courts. Given this scenario, it is imperative for regional public ministries to update their tools for both the early detection and technical verification of these frauds. Institutional success will depend on the acquisition of forensic analysis technologies, continuous training to prevent falling behind criminal advancements, and leadership in digital literacy within the justice system. Ultimately, a profound understanding of AI will play a predominant role in ensuring the efficacy of criminal prosecution and the evaluation of evidence within this new digital paradigm.

Litigation and Judicial Activity Analysis

AI tools are transforming how legal professionals access, systematize, interpret, and apply legal information to specific cases. Implementing platforms designed to summarize case law, analyze binding institutional criteria, or identify decision-making patterns allows for a more agile and

consistent integration of the applicable legal framework. In aggregate, these tools enhance the quality of decisions and institutional consistency. The following sections address the primary advantages and risks associated with these applications.

Potential Advantages

Prosecution offices act as institutional litigants before courts of law through the exercise of public criminal action. In this role, they seek the optimal resolution to criminal conflicts, operating within a margin of discretion that varies according to case specificities and national or state regulations. In practice, this may involve dismissing a case, promoting alternative dispute resolution, reaching a reparatory agreement, or seeking a conviction.

AI can significantly contribute to these decisions by improving the quality and scope of available information analysis. One example is [Fiscal Heredia](#), developed by the Public Ministry of Chile. This system utilizes AI to process large volumes of documents, extract relevant entities (individuals, accounts, emails, telephone numbers), link dispersed information, and enable hybrid searches based on both semantic similarity and textual matches.

Similarly, AI tools facilitate the construction of jurisprudential databases that integrate content automatically, providing prosecutors with rapid and practical access to judicial criteria and relevant resolutions. Once jurisprudential knowledge is established, tools can be developed to extract and process key information for strategic trial decision-making. For instance, classifiers can automatically identify relevant case law, analogous cases, and successful or unsuccessful argumentative lines based on the investigative file. These tools can also integrate with institutional management systems to produce aggregate analyses of judicial outcomes, providing evidence-based feedback to guide prioritization and strategic prosecution.² This

² [Spain announced the use of AI tools](#), including generative AI, within the CENDOJ/CGPJ to draft summaries and extract information from resolutions in its documentary database. [Chile introduced a new jurisprudential search engine](#) that utilizes AI to facilitate the localization of sentences and enhance the search experience within its repository of rulings. [Brazil developed "Victor" for the Supreme Federal Court](#), an AI tool designed to classify and identify legal themes while supporting the management of large caseloads. At the European level, the [CULTEXP project](#) established an AI-powered legal database containing case law and reports from various countries for access and reuse. In the private sector, platforms such as vLex (Vincent), Lexis+ AI, and Thomson Reuters (Westlaw) products integrate AI for search and support functions, including the summary and analysis of jurisprudential repositories.

enables access to pertinent precedents without relying exclusively on manual searches, helping to anticipate evidentiary or argumentative weaknesses and improving the capacity to adapt procedural claims to specific judicial contexts.

Risks

While AI litigation support tools can significantly optimize prosecutorial work, they also pose risks that require careful consideration. These applications do more than facilitate access to case law; by classifying cases, suggesting strategies, or estimating probabilities of success, they influence decisions regarding which cases to pursue and with what level of intensity. Although the final decision remains human, these tools possess the power to influence decision-making by prioritizing certain criteria, conditioning the prosecutor's professional judgment, or introducing biases that affect procedural equity.

Consequently, the data used to train these models, and the performance of algorithmic impact assessments are fundamental to detecting and preventing the reproduction of bias. For example, a system could biasedly recommend denying alternative resolutions to defendants with specific socioeconomic characteristics or those residing in certain areas. This would reinforce structural biases while disregarding the symbolic, preventative, or social value of alternative dispute resolution mechanisms.

Furthermore, the risk of algorithmic opacity remains a concern. If recommendations stem from models whose logic is unintelligible to the user, substantive control over high-stakes prosecutorial decisions is weakened. Therefore, any implementation of AI in the prosecutorial sphere must be accompanied by robust human oversight, traceability, and continuous validation, as well as periodic evaluations of institutional impacts, particularly when these tools affect the exercise of public criminal action.

Risk Assessment of Recidivism and Criminal Offending

The use of AI to predict the probability of recidivism has expanded across the criminal justice sector, with tools now available to courts, prosecution offices, and law enforcement agencies. These systems aim to rationalize key procedural decisions, such as requesting pretrial detention or assessing the risk of domestic and gender-based violence. This section explores the institutional applications of these tools by prosecutors, alongside the technical, ethical, and regulatory challenges they present.

Prosecutorial Function and the Rationalization of Pretrial Measures

Prosecution offices must frequently determine whether to request restrictive measures against defendants. In this context, AI-driven predictive models serve as support tools to evaluate the risk that a defendant's liberty poses to the investigation's success or victim safety—specifically regarding the risk of new offenses. When implemented with appropriate safeguards, oversight, and impact assessments, these tools could objectify the criteria for requesting such measures. This promotes a more proportional application of the law, minimizes decisions based on intuition or stigma, and ensures that pretrial detention remains a measure of last resort, in accordance with international human rights standards³.

Technical Risks: Accuracy, Bias, and Opacity

Despite potential benefits, these tools face significant technical hurdles. As demonstrated by [Kleinberg et al. \(2018\)](#), even highly calibrated models can exhibit bias when applied to diverse populations, particularly if they fail to adjust for base rates of recidivism. Furthermore, trade-offs

³ Different countries have implemented systems of this nature. Spain operates with RisCanvi (for prison classification) and VioGén (for gender-based violence), while the United States has extensively used COMPAS, which has been subject to scrutiny for the reproduction of biases (Angwin and Larson, 2021; Engel et al., 2025)..

between predictive accuracy and algorithmic fairness are frequent. These issues are exacerbated by model opacity. Tools such as COMPAS have faced criticism for their lack of transparency, as they prevent stakeholders from verifying how risk factors are weighted, thereby hindering judicial scrutiny and the right to a defense.

A central risk of predictive tools is the reliance on "dirty data." When models are trained on historical records produced by selective policing or prosecutorial practices—which in some contexts are legally or constitutionally questionable (e.g., mass stops based on racial profiling)—the algorithm learns patterns of biased surveillance rather than actual criminality. In practice, this functions as a self-fulfilling prophecy, reinforcing the overrepresentation of minorities and heavily monitored neighborhoods. This bias is further compounded because the selection of which crimes to predict is not socially neutral. For example, prioritizing property crimes over white-collar or corruption offenses legitimizes a criminal-policy selectivity that often lacks evidence-based planning or proportionality ([Castets-Renard 2021](#)).

An additional concern is the "feedback loop effect." When a system labels specific individuals or neighborhoods as "high risk," it tends to direct increased patrols and police interactions toward them. This heightened contact generates new administrative records (arrests, identity checks, or reports), which are fed back into the algorithm as "evidence" of higher risk. The result is a cycle that confirms its own assumptions, reflecting an increase in detection and data volume rather than an actual change in criminal behavior, thus consolidating historical inequalities ([Brayne & Christin 2021](#)). Similarly, the "selective labels problem" in training data has been identified; unobserved variables—such as a defendant's demeanor during a hearing—can influence detention decisions. The omission of these variables can lead to inaccurate or biased system recommendations ([Kleinberg et al. 2018](#)).

Institutional Risks: Democratic Control, Displacement of Discretion, and Professional Resistance

The substitution of or interference with democratically produced norms by programming code—such as the necessity of caution or the standards of conviction required for a sentencing—remains deeply problematic. Both are fundamental components of the rule of law. Neither the private firms nor the programmers designing these systems are subject to the same normative, judicial, or civic oversight that governs the creation and application of law. Consequently, there is a risk of delegating fundamental democratic functions to algorithms, thereby influencing coercive decisions without equivalent guarantees of public deliberation, contestability, or accountability. For these reasons, their use must be regulated under the strictest standards.

Furthermore, the use of AI to estimate criminal risk may transform the exercise of prosecutorial discretion. These technologies do not replace human decision-making; instead, they displace it toward less visible areas, such as algorithmic design or the pre-selection of variables, which diminishes both internal and external accountability (Brayne & Christin 2021). This reconfiguration of decisional authority has generated resistance among criminal justice professionals, who report concerns regarding deprofessionalization, intensified managerial control, and the imposition of technical standards foreign to legal logic. Such dynamics may lead to informal resistance strategies, including the selective use of tools or, in extreme cases, the manipulation of input data, ultimately compromising the quality, consistency, and democratic legitimacy of the justice system.

Regulatory Recommendations, Impact Assessment, and Democratic Control

In response to these risks, several international proposals advocate for the development of robust and integrated regulatory frameworks, as the mere declaration of principles does not guarantee adequate outcomes. Consequently, it is necessary to establish mechanisms that:

- Evaluate differentiated impacts on vulnerable populations.
- Ensure that models are auditable, explainable, and modifiable.
- Guarantee the ability to challenge and contest algorithmic decisions.

- Limit or prohibit the use of these tools in contexts where the potential harm to fundamental rights is disproportionate or excessively high.

This approach entails not only monitoring models after their design but also engaging in institutional reflection to determine whether their use is appropriate in the first place, under what conditions, and with which safeguards. Accountability throughout the entire algorithmic lifecycle—from conception to application—is indispensable for preserving the democratic legitimacy of the criminal justice system.

Strengthening Compliance with Institutional Guidelines

Unlike the judiciary, high-ranking authorities within public prosecution offices typically possess the power to issue mandatory institutional guidelines. These directives establish minimum investigative steps, enabling conditions for specific actions—such as validation from a jurisdictional lead prosecutor or a specialized unit head—and instructions for interacting with victims and witnesses, among other essential investigative and prosecutorial protocols.

While such guidelines are vital for standardizing prosecutorial work, their effective application in high-caseload environments faces multiple obstacles. The broad range of criminal offenses generates an equally high number of specific instructions. Coupled with the sheer volume of cases assigned to the average prosecutor, the format of these guidelines—often found in extensive or fragmented technical reports—frequently renders their systematic consultation and case-specific application unfeasible in practice.

These limitations have severe consequences. Failure to perform mandatory investigative steps intended to guarantee minimum standards compromises institutional efficacy and reduces the probability of obtaining convictions, a sustained concern for regional public ministries. Furthermore, organizational oversight is hindered; prosecution offices often lack effective mechanisms to monitor and audit compliance given the massive volume of records, the length of investigative files, and the limited personnel dedicated to internal control.

Against this backdrop, AI can enhance the compliance and effectiveness of institutional guidelines through three complementary approaches.

First, AI systems can facilitate the systematization, verification, and updating of institutional guidelines. These tools can extract, classify, and organize instructions contained in various institutional documents to consolidate dynamic and accessible knowledge bases. Such databases allow for the identification of current guidelines for each criminal category and the detection of contradictions between directives issued by different units or at different times, thereby streamlining regulatory updates and reducing information dispersion.

Second, AI can be implemented as a proactive interface designed to inform the prosecutor in charge of a case regarding applicable mandatory guidelines. By integrating case management systems, AI can automatically detect the specific offense and suggest the investigative steps or authorizations required by current institutional protocols. This enables the generation of alerts, reminders, or automated checklists that assist the prosecutor in investigative planning, reducing the risk of omissions caused by lack of knowledge or excessive workload.

Third, AI technologies can serve as tools for ex-post auditing. Through the automated analysis of case files, these systems can contrast the actions actually performed against the applicable mandatory guidelines, identifying potential non-compliance and generating reports for internal oversight. This information can also inform continuous training processes by detecting recurring patterns of non-compliance and designing targeted training programs.

Conclusions

Public prosecution offices are at a decisive juncture, driven by the advancement of digital technologies and the imperative to modernize criminal investigation and prosecution. Establishing guiding principles is an indispensable first step in defining an ethical framework; however, principles alone are insufficient to guarantee the responsible use of automated tools. The effective protection of procedural safeguards and the legitimacy of the justice system depend on the institutional capacity to implement robust AI governance.

This oversight architecture must transcend declarative statements to establish concrete operational mechanisms that regulate the entire lifecycle of these systems. Achieving this requires formalizing strategic decision-making and ensuring internal coherence through mandatory protocols that standardize tool usage. Furthermore, it is imperative to guarantee meaningful human oversight and clear traceability to enable auditing, thereby ensuring accountability and democratic control.

Risk management must be a central element of this new institutional stage. Categorizing usage risks allows for the application of a proportionality criterion essential for distinguishing between administrative applications that enhance efficiency and those that influence criminal investigations. While the automation of bureaucratic tasks offers immediate opportunities for resource optimization, decision-support and predictive tools demand the highest validation standards to prevent bias and error.

Challenges arise not only from internal technological adoption but also from the evolution of criminal behavior. The proliferation of crimes facilitated by generative AI places unprecedented pressure on the capacity of prosecution offices to address current and future threats. The increasing difficulty in distinguishing between authentic content and manipulated material necessitates a reimagining of the criminal procedural system to adapt to these disruptive

advances. Ensuring that the pursuit of truth remains viable in an increasingly complex digital environment will depend on the ability of regional public prosecution offices to adopt and master these technologies.

References

Angwin, J., & Larson, J. (2022). Bias in criminal risk scores is mathematically inevitable, researchers say. En J. Angwin & J. Larson, *Ethics of Data and Analytics* (1.^a ed., p. 3). Auerbach Publications.

Brayne, S., & Christin, A. (2021). Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social Problems*, 68 (3), 608–624. <https://doi.org/10.1093/socpro/spaa004>

Castets-Renard, C. (2021). Human rights and algorithmic impact assessment for predictive policing. En H.-W. Micklitz, O. Pollicino, A. Reichman, A. Simoncini, G. Sartor, & G. De Gregorio (Eds.), *Algorithms, Freedom, and Fundamental Rights* (Part I). Cambridge University Press.

Europol Innovation Lab. (2024). *AI and policing: The benefits and challenges of artificial intelligence for law enforcement*. Publications Office of the European Union. <https://doi.org/10.2813/0321023>

European Union Agency for Law Enforcement Cooperation. (2024). *Decoding the EU's most threatening criminal networks*. Publications Office of the European Union. <https://doi.org/10.2813/811566>

Financial Action Task Force. (2021). *Opportunities and challenges of new technologies for AML/CFT*. FATF. <https://www.fatf-gafi.org/publications/fatfrecommendations/>

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133 (1), 237–293. <https://doi.org/10.1093/qje/qjx032>

Lee, Y., Bradford, B., & Posch, K. (2024). The effectiveness of big data-driven predictive policing: Systematic review. *Policing: A Journal of Policy and Practice*, 127–160. <https://doi.org/10.1093/police/pzae045> [Nota: Ajustado según metadatos del DOI común para este título].

Ligeti, K., Vervaele, J., & Vermeulen, G. (2025). *Artificial intelligence and criminal justice = Intelligence artificielle et justice pénale = Inteligencia artificial y justicia penal*. RIDP libri 12.

Rigano, C. (2018, 8 de octubre). *Using artificial intelligence to address criminal justice needs*. National Institute of Justice. <https://nij.ojp.gov/topics/articles/using-artificial-intelligence-address-criminal-justice-needs>

UNESCO. (2023). *Kit de herramientas sobre la IA y el Estado de derecho*. UNESCO. <http://www.unesco.org/open-access/terms-use-ccbysa-en>

Vargas-Murillo, G. (2024). Code is law: how COMPAS affects the way the judiciary handles the risk of recidivism. *Artificial Intelligence and Law*, 33, 383–404. <https://doi.org/10.1007/s10506-024-09393-x>



Centro de Estudios de Justicia de las Américas
Justice Studies Center of the Americas

Artificial Intelligence in Public Prosecution: Reflections and Guidelines for the Americas



Centro de Estudios de Justicia de las Américas (CEJA)

www.cejamericas.org

