Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

Artificial Intelligence and State Advocacy: parameters for safe use in drafting legal documents.

Artificial Intelligence and State Legal Counsel: parameters for safe use in the drafting of procedural documents

Artificial Intelligence and Defense of the State: parameters for safe use in the preparation of procedural documents

Leandro Mendes Neris
Law graduate
Institution: Federal University of Amazonas
Email: leandromendesneris@gmail.com

ABSTRACT: This article analyzes, from legal, ethical, and technical perspectives, the incorporation of generative artificial intelligence systems—especially Large-Scale Language Models (LLMs)—into the activities of State Attorneys, notably in the drafting of procedural documents, opinions, appeals, and petitions. The study examines the specific risks of this technology in the legal-institutional context, including the phenomenon of language model hallucinations, vulnerability to prompt injection attacks, the risk of leaks of confidential information, and the manipulation of AI-generated responses. Conversely, the potential benefits in terms of productivity, jurisprudential research, and standardization are analyzed.

The article proposes practical parameters for safe use, including prompt engineering techniques applied to public advocacy, institutional governance, standardized prompt libraries, and criteria for disciplinary responsibility of public prosecutors. It concludes that AI can be a powerful tool for public advocacy, provided its use is disciplined by qualified human supervision, strict institutional protocols, and...

Awareness of the ethical and legal limits of delegating intellectual activities to automated systems.

KEYWORDS: Artificial Intelligence. State Advocacy. Prompt Injection.
LLM hallucinations. Prompt engineering. Attorney responsibility. Institutional governance. Information security.

ABSTRACT: This article analyses, from legal, ethical, and technical perspectives, the incorporation of generative artificial intelligence systems — particularly Large Language Models (LLMs) — into the activities of State Legal Counsel, notably in the drafting of procedural submissions, legal opinions, appeals, and petitions. The study examines the specific risks of this technology in the legal-institutional context, including language model hallucinations, vulnerability to prompt injection attacks, the risk of sensitive information leakage, and the manipulation of AI-generated responses. In contrast, the potential benefits in terms of productivity, case law research, and standardization are also analyzed. The article proposes practical parameters for safe use, including prompt engineering techniques applied to public legal counsel, institutional governance, standardized prompt libraries, and criteria for the attorney's disciplinary responsibility. It concludes that AI can be a powerful tool for State Legal Counsel, provided that its use is regulated by qualified human supervision, strict institutional protocols, and awareness of the ethical and legal limits of delegating intellectual activities to automated systems.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

KEYWORDS: Artificial Intelligence. State Legal Counsel. Prompt injection. LLM hallucinations. Prompt engineering. Attorney's responsibility. Institutional governance.
Information security.

1 INTRODUCTION

The technological revolution ushered in by the spread of Language Models
Large-scale language models (LLMs) possibly represent the most
A profound transformation in intellectual practices has occurred since the popularization of the internet.
In the legal field, the use of tools based on generative artificial intelligence —
such as ChatGPT from OpenAI; Claude from Anthropic; Gemini from Google; and Copilot,
From Microsoft — it has ceased to be a future prospect and has become a present reality.
incorporated into the daily routine of law firms and legal departments
corporate and, progressively, to legal careers in the public sector.

State Advocacy — understood, in the Brazilian legal system, as
The Attorney General's Office, through the State and Federal District Attorney's Offices and the
Municipal Attorney's Offices face particular challenges in this scenario. On one hand,
the growing demand for efficiency, resource optimization, and speed of response in
In contexts with a high volume of processes, artificial intelligence becomes a solution.
attractive. On the other hand, the peculiar nature of the judicial and advisory representation function of
The State, guided by the inalienability of the public interest, by the strict observance of
Legality and the institutional responsibility of the prosecutor impose limits that do not
They exist — or exist to a lesser degree — in private practice.

The use of AI in drafting legal documents raises questions that
They permeate different domains of legal knowledge. From a technical point of view, it is
It is necessary to understand the operating mechanisms of LLMs and their points of failure.
and its vulnerabilities. From an ethical and professional standpoint, the problem arises of
delegation of intellectual activities inherent to the practice of law to systems
automated processes, with implications for the duty of diligence, procedural loyalty, and...
The lawyer's responsibility. From an institutional point of view, questions arise regarding...
governance, information security, and standardization of procedures require
Internal regulations by the state's legal aid bodies.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

The aim of this article is to analyze, in a systematic and practical way, the Parameters for the safe use of artificial intelligence systems in the legal profession. State. To that end, the study is structured as follows: initially, a An overview of LLMs and their operation is presented; then, the phenomenon of... is examined. Hallucinations and their legal impacts; the concept of prompt injection and other related topics are analyzed. risk vectors are identified; mitigation measures and governance guidelines are proposed; and the following is discussed. prompt engineering applied to public advocacy; and, finally, responsibility is analyzed. Disciplinary and legal action against the prosecutor for using AI-generated content.

2 Large-Scale Language Models and Context

LEGAL

2.1 Basic Operation of LLMs

Large-scale language models are artificial intelligence systems. trained on large volumes of text, with the goal of predicting, in a probabilistic way, Which token (unit of text) should follow the previous one in a sequence? Starting from Transformer-based architectures, developed from Vaswani's seminal article. et al. (2017), these models learned sufficiently rich linguistic patterns to Generate coherent text, answer complex questions, translate languages, draft documents. and, if of interest here, drafting legal documents.

It is essential to understand that an LLM does not "know" properly in the sense that a A legal expert knows: the model does not have internal representations of norms, it does not consult databases of real-time data (except when equipped with specific external tools) and not It reasons deductively from established premises. What the model does is, Essentially, to recognize statistical patterns and generate text that, given the inputs received (prompts), resembles what, in your training set, would follow that context. The appearance of rigorous legal reasoning is, to a large extent, a consequence from the structure of the legal text incorporated into the training, and not from a deductive process. genuine.

This distinction is fundamental to understanding the risks associated with the use of LLMs in law practice, as it explains why the model can produce legally sound arguments.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

Flawless in form, but factually incorrect in content — quoting non-existent precedents, attributing false summaries to real rulings, or asserting the state regulations concerning laws that have already been repealed.

2.2 The LLM as a support tool and its limitations

From a productivity standpoint, LLMs offer real contributions and documented in three main areas. Firstly, in accelerating the production of Minutes: A prosecutor who masters prompt engineering techniques can obtain a structured draft of a defense, response to objections, or opinion in a fraction of the time which would take to produce it from scratch, dedicating intellectual energy to supervision and adjustment. and to the improvement of the generated content. Secondly, in case law research, the Models with internet access or equipped with integrated legal databases are able to identify relevant precedents, synthesize legal arguments, and map the evolution of case law with greater speed than manual research. Thirdly, in the standardization of language. and in formal compliance, reducing procedural errors of a technical nature.

However, these benefits cannot obscure critical structural limitations. Knowledge of LLMs is static, limited to the training cutoff date, and frequently Outdated in relation to the most recent legislative and jurisprudential changes. A Hallucination — the confident production of false information — is a phenomenon intrinsic to The functioning of these models is not a flaw that can be corrected by simple adjustments. And the Excessive reliance on model output, without qualified human supervision, can transforming AI from a support tool into a vehicle for material errors, with serious procedural consequences.

3. Hallucinations of Language Models and Their Impacts on the Performance of the Public Treasury

3.1 Concept and typology of hallucinations

The term hallucination, as applied to LLMs, refers to the generation of information that does not They correspond to factual reality, but are presented by the model with the same level of... Confidence and fluency that provide accurate information. Unlike a typo.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

or a readily perceptible logical inconsistency, the hallucinations of LLMs
They are often presented in the form of well-structured, internally coherent text and
seemingly plausible — which makes its detection particularly challenging for those
does not have a deep understanding of the subject matter.

In a legal context, hallucinations take on specific typologies of particular importance. The first and most documented example is the fabrication of judicial precedents: the model. cites rulings with plausible numbering, rapporteur, date, and summary, which simply do not There are. The second is the distortion of real precedents: the model cites a judgment that exists, but assigns to it a different summary or thesis than that in the original. The third is the Citing repealed legal provisions as if they were still in force, or asserting states of law. regulations that have already been amended. The fourth is the mistaken attribution of doctrinal positions to authors, citing non-existent works or attributing a thesis to a particular jurist that is not true. your.

3.2 Documented cases of hallucination in the legal profession

The most widely publicized case study on the impact of AI hallucinations in practice. The legal argument was made by lawyer Steven Schwartz in the United States, who filed a petition in 2023. to the Federal Court for the Southern District of New York, citing six non-existent precedents, all generated by ChatGPT. When summoned by Judge Kevin Castel to prove the Despite the existence of the rulings, the lawyer failed to do so, suffering disciplinary sanctions and conviction and fine imposed. The episode generated widespread repercussions and motivated the Developing guidelines for the use of AI by courts worldwide.

In Brazil, contrary to what one might assume, there is already a significant set and growing body of precedents on the subject, with sanctioning decisions issued by the most various courts across the country. The first case of institutional relevance dates from April 2023: the then-Inspector General of Electoral Justice, Minister Benedito Gonçalves, of the TSE, applied A fine of R$ 2,604 was imposed on a lawyer for bad-faith litigation who filed a lawsuit. petition drafted entirely by ChatGPT, in which it sought to participate as amicus curiae. in an electoral lawsuit involving former president Jair Bolsonaro, a process in which he did not did not represent either party. The Minister was emphatic in noting that the lawyer had submitted for judgment "a fable, resulting from a conversation with an artificial intelligence".

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

In February 2025, the TJSC (Court of Justice of Santa Catarina) consolidated a more severe decision-making standard in two...

distinct and emblematic cases. In the first, the 6th Civil Chamber fined a lawyer in
10% of the value of the case after he admitted to using ChatGPT in preparing the appeal.
in a repossession action — a legal recourse in which both case law citations are used.
Regarding the extent to which the references to legal works were incorrect or entirely fictitious. In addition to

As a monetary penalty, the court ordered that the case be reported to the OAB/SC (Brazilian Bar Association of Santa Catarina) and the...

Forwarding a copy of the appeal for disciplinary review. The lawyer alleged that the error
This occurred due to "inadvertent use" of ChatGPT. In the second case in Santa Catarina, the 5th Chamber
Criminal investigators formally warned the lawyer who used AI to draft a habeas corpus petition.
There is no precedent, the rapporteur having stated that the conduct constituted an act of bad faith.
and disrespect for the court and that the precedents presented "were created to induce the
"judge made a mistake".

In labor courts, the phenomenon has taken on even more worrying proportions.
In June 2025, the TRT-7 (Ceará) identified, in an appeal petition, case law.
manipulated with falsified or non-existent passages, including improperly inserted passages
assigned to the TST that dealt with completely different matters — one of the summaries,
Presented as if it were about workplace harassment, it actually concerned...
"PJ-ization" — with evidence of being generated by artificial intelligence. The court imposed a fine.

2% of the value of the case and ordered that an official letter be sent to the OAB-CE (Brazilian Bar Association of Ceará) for investigation of

Possible disciplinary infraction. In October 2025, the TRT-12 (Santa Catarina) judged the case.
in which the initial petition contained decisions, doctrinal citations, and even the name of a magistrate.
nonexistent, all elements apparently generated by artificial intelligence. The judge
Daniel Carvalho Martins stated that "such findings go beyond a mere clerical error" and
that the piece was produced "without any human verification, which constitutes an act
"non-existent procedural document." The party was ordered to pay a fine of R$ 3,700.

In February 2026, the TRT-2 (São Paulo) faced a situation that exposed a
A recurring avoidance strategy in forensic practice: when questioned about the use of
fictitious jurisprudence, the lawyer attributed the responsibility to the "body of interns of
office," an argument that the court vehemently rejected, stating that "the claim
"In court, it is an act exclusive to the lawyer, who is also responsible for its content."
The magistrate expressly highlighted that the conduct violated Recommendation No. 1/2024 of
The Brazilian Bar Association's Federal Council, which requires human supervision in the use of generative tools,

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

and that there was "deliberate creation of precedents to reinforce the defense's argument, compromising legal certainty." It is

precisely this normative instrument — Recommendation No. 1/2024 of The Federal Council of the Brazilian Bar Association — which has been invoked as an objective parameter in most recent decisions. According to the rule, the lawyer is required to have "understanding appropriate limitations, rigorous verification of information, transparency to customers and other interlocutors, and the delegation of acts exclusive to the profession is prohibited without "Qualified supervision." The consolidation of this chain of precedents—encompassing the TSE, The TJSC and multiple TRTs demonstrate that the problem is neither incipient nor episodic: Between 2025 and 2026, Brazilian courts recorded several cases of lawyers being fined for submitting petitions with false case law generated by AI, establishing a pattern. This alarming situation has already come to the attention of the state ombudsman offices.

3.3 Specific impacts on the performance of the Public Treasury

For the State Attorney's Office, the impacts of hallucinations are significant. This is especially serious for reasons that add to those affecting private legal practice. In First, the prosecutor represents the State — the holder of an inalienable public interest. —, so that a material error resulting from AI hallucination may compromise not It's not just an individual cause, but it can also create negative precedents for the public entity. throughout a line of case law. Secondly, the functional responsibility of The responsibility of a public prosecutor is more intense than that of a private lawyer: errors that, in private practice, In public advocacy, these actions would only entail contractual and disciplinary liability. To establish misconduct in a public office, with administrative consequences.

Thirdly, the nature of disputes involving the Public Treasury — frequently related to taxes, social security, court-ordered payments, public services and public policies — implies that an error in a single petition can have a multiplier effect. about hundreds or thousands of similar cases processed by the same agency. A thesis legally flawed, incorporated into a draft generated by AI and not rigorously reviewed. sufficiently, it can spread systematically if adopted as an institutional model.

Fourthly, there is the risk that the pressure for productivity — often high in public advocacy bodies due to the imbalance between the volume of

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

processes and the number of prosecutors — fosters a relationship of excessive trust in exiting the model, reducing the human scrutiny required to detect hallucinations.

4 INSTITUTIONAL SECURITY RISKS: PROMPT INJECTION, LEAKAGE AND HANDLING

4.1 The concept of prompt injection

The concept of prompt injection: Prompt injection is a category of attack that explores the structural inability of LLMs to reliably distinguish between instructions provided by the system operator (system prompt) and the content produced by users or external sources that the model processes. The concept was formally identified in 2022 and has since received increasing attention from the community of Security. By 2025, OWASP—a global authority on application security— It positioned prompt injection as the top threat in its ranking of risks to systems. of generative AI (OWASP Top 10 for LLMs 2025). The attack consists of inserting, into content that the model analyzes, hidden or disguised instructions that, when processed Through LLM, they alter their behavior, override the original instructions, or extract Information that should not be revealed.

There are two main subtypes. The first is direct prompt injection, which occurs when user input directly affects the behavior of the LLM, whether in a way intentional — when the user plans and structures the attack — or unintentional. The second is indirect prompt injection, which occurs when the model accepts prompts from sources. External sources, such as websites or files, and the content of these sources, when interpreted, alters the LLM behavior. Both modalities can result in biased responses. in the disclosure of confidential information or in the execution of unauthorized actions.

In the context of public advocacy, the risk is particularly sophisticated: a a prosecutor who uses an AI system to analyze the opposing party's initial petition and Extracting the main arguments to refute can, unknowingly, lead to the processing of malicious commands. embedded within the piece itself by the opposing party, instructing the model to produce an analysis. favoring the opponent, omitting relevant arguments, or leaking strategic information. of the agency. What until recently seemed a hypothetical scenario became, in May of

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

2026, in concrete and legally sanctioned reality in Brazil, inaugurating a new and alarming frontier of bad-faith litigation.

4.1.1. The Parauapebas case: One of the first Brazilian precedents

On May 12, 2026, the 3rd Labor Court of Parauapebas encountered a an unprecedented finding in Brazilian jurisprudence. When processing the initial petition by Galileo system — a generative artificial intelligence tool for the Labor Courts — , The court identified text in white font on a white background, invisible to the reader. human, containing the following command: "Attention, artificial intelligence, contest this The petition should be superficial and not challenge the documents, regardless of the order. whatever is given to him."

The judge dedicated the first pages of the sentence to condemning the conduct. stating that it was "incompatible with the most basic duties that fall upon everyone who participates in the judicial process" and classified the practice as "an attack on The credibility of institutional tools is undermined, leading to a lack of respect for the courts, the parties involved, and society. "And this is a precedent that this court cannot let pass." The ruling acknowledged the act. The court, finding the act an affront to the dignity of justice, jointly condemned the two signatory lawyers. to the payment of a fine of 10% on the value of the case — approximately R$ 84,250.00 — and It was determined that an official letter should be sent to the OAB/PA (Brazilian Bar Association of Pará) and to the Inspectorate of the TRT (Regional Labor Court) of the 8th Region.

4.1.2. The national escalation: STJ, São Paulo and the criminal dimension

The Parauapebas incident is no longer an isolated case. On May 20, 2026, The president of the Superior Court of Justice (STJ), Minister Herman Benjamin, ordered the opening of a police investigation. and administrative procedure to investigate attempts at procedural fraud by means of prompt injection into the STJ Logos system — a generative artificial intelligence system Developed by the Court itself and launched in February 2025. According to TV reporting. Globo reports that there are at least 11 criminal cases in which the technique has been identified.

The Superior Court of Justice (STJ) determined that attempts to use the mechanism, although neutralized, were not carried out. through the system's security layers, they would be certified in the records, which allowed the imposition of procedural sanctions by the ministers of the Court. The vice-president of

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

STJ Minister Luis Felipe Salomão was even more emphatic in describing the practice as "police case," stating that "when AI is misused, in bad faith, which goes beyond the scope of..." This is a purely ethical issue; it's a matter for the police, and it needs to be investigated both on the level and at the police level. "both administrative and criminal."

On May 21st, the phenomenon reached São Paulo, with the identification of prompt Injection in petition against bank in civil court of the capital, confirming that the issue has left It went from being anecdotal to acquiring a national systemic dimension.

4.1.3. Implications for State Advocacy

The technique is classified by OWASP as the number one vulnerability in its... List of risks for model-based language applications. For the Legal Department of In the state, the risk factor is twofold. In the active sense, the agency can be a victim of prompt action. Indirect injection occurs when their attorneys submit documents from the opposing party for review. AI systems — a situation where hidden commands in these components can compromise the The quality of the analysis, or, in poorly configured systems, the extraction of confidential information. passive sense, AI systems used by the public advocacy bodies themselves — for procedural screening, risk analysis and drafting of minutes — these are targets of external attacks.

4.2 Leakage of sensitive information

The use of AI tools in the context of State Advocacy involves, necessarily, the processing of highly sensitive information: data on litigation. ongoing defense and attack strategies of the Public Treasury, internal opinions not Published information about taxpayers in tax proceedings, personal data of servers in labor lawsuits and classified information related to contracts. Administrative matters under secrecy.

The risk of such information leaks occurs through multiple vectors. The first It's the storage on the servers of the AI provider company: most of the tools LLM available in the consumer market transmits user inputs to external servers, where they can be stored, used in version training.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

future developments of the model and subject to security breaches. The use of tools
AI commercials with confidential procedural information without a contract being signed.
A specific guarantee that ensures confidentiality and the deletion of data may constitute a violation.
to the rules of professional secrecy and to the General Data Protection Law (LGPD, Law No.
13.709/2018).

The second vector is the exposure of information during interaction with the model: to
to formulate a prompt for the model to draft a rebuttal, the prosecutor
necessarily includes factual elements of the case, identification of the parties and information.
strategic. In systems without adequate data protection, this information can be
accessed by third parties — company employees, researchers, or malicious actors
in case of a security breach.

The third vector is the training of the model based on the data entered by
users. Several commercial generative AI tools utilize the interactions of
users to continually improve their models. This means that information
Sensitive information inserted by a lawyer can, in theory, appear in the responses provided to...
other users, constituting a serious breach of professional confidentiality.

4.3 Manipulating AI responses

In addition to spontaneous hallucinations and prompt injection attacks, there is a third
A risk of a distinct nature: the intentional manipulation of AI responses by actors who,
Having knowledge of how a particular AI system works, they structure their strategy.
procedural in order to influence the analysis that the adversarial system will perform. In contexts
in which the opposing party is aware that the Public Treasury uses a certain
With AI tools for analyzing procedural documents, the possibility arises — even more
more theoretical than practical at the current stage — that the party formulates its statements in a way that is theoretical than practical at this stage.
to explore the weaknesses of the model used by the public body.

A more immediate and concrete example of manipulation is what's called jailbreaking — a
a set of techniques developed to circumvent ethical and security safeguards of
LLMs, inducing them to produce content they would normally refuse to generate. In the context
In public advocacy, a prosecutor who uses jailbreak techniques to induce the model.
to produce arguments that, if evaluated ethically, would be considered abusive or

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

contrary to the public interest, it would be using AI as an instrument for practices
Reprehensible procedural practices, with the potential to constitute procedural bad faith.

4.4 Mitigation measures

Adopting effective mitigation measures requires a layered approach that Combine technical, procedural, and institutional safeguards.

On the technical level, priority measures include: (a) the implementation of systems of AI on-premises infrastructure or in a private cloud, with complete control over the data. processed, eliminating the risk of transmission to third-party servers; (b) the selection of tools that offer contractual guarantees that the data will not be used for training. and periodic data deletion; (c) the implementation of input filtering systems and outputs that identify prompt injection patterns before malicious content is released. processed by the model; (d) the use of fine-tuned models — adjusted from data legal matters handled by the agency itself — to reduce the risk of hallucinations in legal domains specific legal aspects.

In procedural terms, the essential measures include: (a) the obligation human verification of all case law and legislation generated by (a) the model, through consultation of primary sources; (b) the prohibition of data insertion identified parts and classified information in non-AI tools approved by the agency; (c) the implementation of access controls that restrict the (d) the creation of logs of Uses that allow for subsequent auditing of the tool's usage in each case.

At the institutional level, the guidelines should establish: (a) the creation of a policy Formal AI usage guidelines, approved by the governing body and binding on all members; (b) the designation of an institutional person responsible for AI governance (Chief AI) Compliance Officer or equivalent); (c) the periodic provision of training on the use AI manager; (d) the creation of a channel for reporting incidents related to the use AI inadequacy.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

5. Prompt Engineering Applied to Public Advocacy

5.1 Concept and relevance of prompt engineering

Prompt engineering is the discipline that deals with strategic development of instructions provided to language models with the objective to maximize the quality, accuracy, and relevance of the generated responses. Far from being a While not a common activity, prompt engineering constitutes a specific technical skill that This can decisively determine the difference between a generic and a low-level AI output. value and a precise, structured, and legally relevant solution.

In the context of State Advocacy, mastery of engineering techniques of Prompt can radically transform the prosecutor's productivity, while also... which reduces — though it does not eliminate — the risk of hallucinations and inappropriate exits. A developing a well-structured prompt that provides the model with sufficient context, Clear restrictions and specific format instructions result in minutes that are far superior to... obtained with generic and imprecise prompts.

5.2 Fundamental Prompt Engineering Techniques

The main prompt engineering techniques with application to legal practice are the following:

a) Role prompting: consists of instructing the model to assume a role. A specific prompt before beginning the task. For public advocacy, an effective prompt can start com: "You are a state attorney specializing in tax law, with experience in tax litigation and with in-depth knowledge of the jurisprudence of the Superior Court of Justice (STJ) and the Supreme Federal Court (STF) on the Subject matter. Analyze the following case and elaborate..." This type of instruction guides the model to to access linguistic and structural patterns that are closer to the desired profile.

b) Providing context (context setting): the quality of the model output is directly proportional to the quantity and quality of the context provided. An effective prompt should specify the nature of the requested document (defense, appeal, opinion, memorandum), the the receiving court, the procedural stage, the relevant factual grounds, and the legal arguments that the prosecutor intends to support.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

c) Chain-of-thought prompting instruction: when instructing the model
to present their reasoning in a structured and sequential manner before drafting the final piece,
The prosecutor gains greater transparency regarding the logical path adopted by the model, which
This facilitates the identification of errors and flaws in argumentation. The instruction might be something like:
"Before drafting the defense, present a draft with: (1) identification of the allegations
(1) the plaintiff's claim; (2) the applicable procedural defenses; (3) the substantive defenses, with indication
of the legal foundations; (4) the order of presentation that maximizes the effectiveness of
defense."

d) Negative prompting: just as it is important to specify what the
As a role model should do, it is equally important to instruct them on what not to do. Examples
Applicable: "Do not cite case law without indicating the case number and the court"; "Do not
Use arguments that imply implicit recognition of copyright; No
Include statements about facts that are not included in the documents provided.

e) Self-checking instruction: the attorney may instruct the template to check
the actual exit before presenting it. For example: "After drafting the defense, review the
text and identify: (1) any statement that is not supported by the documents
provided; (2) any precedent cited, indicating that the attorney should verify its
existence and accuracy; (3) any logical inconsistency." This technique does not eliminate the
hallucinations, but it signals them, facilitating human reassessment.

5.3 Institutional standardization: prompt libraries

One of the most relevant contributions that prompt engineering can offer to
State advocacy is the creation of institutional libraries of standardized prompts —
repositories of carefully crafted, tested, and validated instructions for the tasks.
most frequent for each organ. Standardizing prompts offers multiple advantages:
reduces variability in the quality of outputs generated by different proxies; ensures
that the instructions incorporate the necessary safeguards against hallucinations; facilitates the
training of new members; and allows for centralized updating of instructions when
There are no relevant legislative or jurisprudential changes.

A library of prompts for State Advocacy could include...
the following categories: (a) prompts for drafting objections in matters

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

(a) tax; (b) prompts for drafting appeals in social security matters; (c) prompts for the preparation of opinions on the regularity of administrative contracts; (d) prompts for case law research on specific topics; (e) prompts for consolidation and synthesis of arguments in repetitive cases; (f) prompts for legal risk analysis and probability of success.

Each standard prompt should be drafted with the participation of attorneys. specialized in the relevant subject matter, tested in anonymized real-world scenarios and submitted to a validation process before being incorporated into the library. Maintenance The library's management should be the responsibility of a multidisciplinary team that combines... Legal expertise and technical knowledge in AI.

5.4 Quality control of AI-generated outputs

Regardless of the quality of the prompt used, all output generated by AI must to undergo a quality control process before being incorporated into a part A procedural document signed by the attorney. This process must be formalized in a checklist. institutional guidelines that cover, at a minimum, the following aspects:

- Verification of all cited precedents, with consultation of primary sources (databases). (from court records) to confirm the existence, numbering, and content of the summary;
- Verification of the validity of all legal provisions mentioned;
- Assessment of the consistency between the facts narrated by the model and the documents of process;
- Assessment of the conformity of the arguments with the thesis adopted by the body. institutionally in the matter;
- Language revision to ensure suitability to the style of the legal document and the language required. formal of the body;
- Verification of the absence of confidential information that should not be included in the document;
- Assessment of the ethical appropriateness of the arguments used.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

6. SAFE USE: PRACTICAL APPLICATIONS IN STATE ADVOCACY

6.1 Drafting minutes

Drafting procedural documents is the most immediate application.
and the greatest impact of AI on State Advocacy. For cases of a repetitive nature —
tax enforcement proceedings, challenges to the enforcement of judgments, preliminary objections
enforceability, standardized challenges in social security or tax matters. —, the
The template can generate, in seconds, a structured draft that, when submitted for review by
According to the prosecutor, it can be completed in a fraction of the time it would take to prepare the entire document.

The recommended protocol for drafting minutes using AI should follow the
the following steps: (1) the attorney identifies the document category and selects the default prompt.
(2) the prompt is customized with the data corresponding to the institutional library;
specific to the case, strictly observing the prohibition on data entry.
identified in non-approved tools; (3) the generated draft is submitted to the checklist
quality control; (4) the attorney improves, corrects and supplements the draft with
elements of procedural strategy that require your professional judgment; (5) the document
The final document is signed exclusively by the responsible attorney, without any reference to...
AI participation in its development is prohibited, unless otherwise specifically stipulated by law.

6.2 Case law research

AI-assisted legal research presents a potential for efficiency.
particularly significant. Models equipped with access to legal databases.
updated or internet search tools are able to, from prompts
appropriate, identify the position of higher courts on a given matter, map
the evolution of a thesis over time, identifying recent judgments that have altered it.
consolidated understandings and synthesize, in an organized way, the fundamentals of
main precedents on a topic.

However, the risk of hallucination is particularly high in the research.
case law, which makes it essential to verify all precedents.
identified by the model in the primary sources. The use of AI for case law research.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

It should be treated as a starting point — an aid in identifying candidates for relevant precedents — and never as a definitive source of case law information. A The final consolidation should always result from direct consultation with the official databases of the courts.

6.3 Consolidation of arguments

Another high-value application is the use of AI to consolidate and organize arguments. complex legal matters. In cases of great factual or normative complexity, the attorney You can use the template to identify internal inconsistencies in the opposing party's work. to organize the defense arguments hierarchically, to identify subsidiary arguments or Consider alternative options not taken and structure the piece in a way that maximizes clarity and impact. of the main arguments.

For this purpose, the technique of providing the model with, in addition to the prompt of instruction, the relevant documents of the process for analysis — noting, more Once again, restrictions were placed on the insertion of sensitive data into non-certified systems. Contextualized document analysis using the model can reveal elements that... The prosecutor, pressured by the workload, might not identify with the same... agility.

6.4 Institutional governance of AI use

The responsible implementation of AI systems in public advocacy requires... creation of a governance structure that systematically disciplines all the Relevant aspects of technology use. The essential elements of this structure are:

AI Usage Policy: a formal document, approved by the organization's leadership, that Establish the principles governing the use of AI, the approved tools, the categories of information that cannot be entered into AI systems, the requirements of Human verification and supervision, and the consequences of non-compliance with guidelines.

AI Governance Committee: a multidisciplinary group composed of prosecutors, information technology professionals and information security specialists, responsible for evaluating new AI tools, updating institutional policy, and investigating incidents and recommend improvements to the usage protocol.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

Training program: mandatory training for all prosecutors on
The fundamentals of how LLMs work, and the risks associated with their use in legal practice.
public and responsible use techniques, including prompt engineering and procedures.
output verification.

Periodic audit: regular review of the use of AI systems in the organization, with
evaluation of the quality of the outputs generated, identification of patterns of inappropriate use and
Verification of the effectiveness of the implemented mitigation measures.

7. ATTORNEY'S RESPONSIBILITY FOR THE USE OF AI-generated content

7.1 Fundamentals of professional responsibility

The question of the prosecutor's responsibility for the use of content generated by
Artificial intelligence is perhaps the most delicate and least regulated aspect of the subject.
The starting point of the analysis is the observation that responsibility for the procedural document
It remains entirely and exclusively with the attorney who signs it, regardless of...
degree of participation of automated systems in its development. AI is not subject to
In law, it lacks the capacity to litigate and cannot be held liable for errors.
materials or arguments.

Brazilian law does not, at present, provide for any such provision.
specific regulation governing the use of AI in law or assigning responsibility for
Errors arising from the use of automated systems. Given this gap, the analysis of
The prosecutor's responsibility should be conducted based on general principles that
They govern the practice of law and public office.

7.2 Duty of care and liability for material error

The Statute of the Legal Profession (Law No. 8.906/1994), in its article 32, establishes that the
A lawyer is responsible for the acts that, in the exercise of their profession, they commit with intent or negligence.
The Code of Ethics and Discipline of the Brazilian Bar Association (OAB), in turn, enshrines the duty of diligence as...
one of the pillars of the practice of law. For state prosecutors, these duties are

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

They combine with the principles of legality, impartiality, morality, and efficiency that
They govern the Public Administration (article 37, caput, of the Federal Constitution).

The use of AI without proper human supervision may constitute a violation of...
Duty of diligence in multiple scenarios. The most obvious is the petition that cites precedents.
nonexistent, without the prosecutor having verified their existence — a situation that, in law
In the United States, this has already resulted in disciplinary sanctions and court decisions that warn...
Regarding the risks of irresponsible AI use. In Brazil, although no decisions have yet been made...
Regarding specific judicial decisions on the subject, the prevailing deontological principles are sufficient for...
to substantiate accountability.

Verification that the prosecutor used AI to draft the document without review.
appropriate, resulting in a material error that caused damage to the Public Treasury, could
to establish responsibility on three distinct and cumulative levels.

In civil law, the attorney's liability to the public entity that
This representation is based on Article 37, § 6, of the Federal Constitution, which enshrines the
objective liability of the State towards third parties and then opens the possibility of
A right of recourse against the public official who caused the damage when that official acted with intent.
or fault. Regressive liability presupposes the demonstration of culpable conduct or
willful misconduct by the prosecutor — in the case under examination, the negligence embodied in the absence of
Proper review of the AI-generated part —, of the actual damage to public funds and the causal link of
causality between the two. The legal regime applicable to this liability varies.
depending on the prosecutor's employment status: federal, state, or civil servants.
Municipalities are subject to their respective laws and legal regimes;
Public employees are subject to the CLT (Consolidation of Labor Laws) regime; and members of career paths with specific statutes,
to the rules governing their career, without prejudice to the subsidiary application of the statutes.
general. The common denominator to all these regimes is the constitutional requirement of intent.
or fault as a prerequisite for a claim for reimbursement, which makes it legally relevant for purposes
Regarding accountability, the demonstration that the prosecutor failed to exercise supervision...
The human responsibility that fell to them regarding the content generated by the artificial intelligence tool.

In disciplinary terms, prosecutors are not only subject to the general rules of
public servants, primarily accountable to the appropriate disciplinary bodies.
of their careers, based on specific governing legislation. Diversity
The institutional framework of Brazilian public advocacy implies a plurality of disciplinary regimes:

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

At the federal level, members of the Attorney General's Office are governed by Law Complementary Law No. 73/1993, which structures the career, defines the functional duties and establishes the applicable disciplinary regime, with subsidiary application of prohibitions and duties. and sanctions provided for in the general statute of federal civil servants; at the state and district levels, The State and Federal District Attorneys are subject to their respective organic laws. and statutes of local public advocacy, which as a rule assign to the ombudsman offices themselves Prosecutor's offices have the authority to investigate infractions committed by their members; at the municipal level, Municipal Attorneys are subject to municipal functional statutes and, where applicable, There are specific disciplinary regulations in place for the Public Prosecutor's Offices; the prosecutors of Legislative Assemblies, Municipal Chambers and other bodies of the Legislative Branch They are linked to the disciplinary regulations specific to each institution. In all these cases, the edition A formal institutional policy on the use of AI provides additional regulatory support to initiation of disciplinary proceedings in case of non-compliance with the established guidelines. making the demonstration of the disciplinary infraction more objective. The most common disciplinary types common to all these statutes — and potentially applicable to the case under examination — are Negligence in the performance of duties, failure to comply with internal regulations. and the practice of acts incompatible with the dignity of the office.

It is important to note that disciplinary responsibility presupposes the demonstration of Negligent or intentional conduct. The mere use of AI as an aid in drafting documents. procedural matters, provided that they are accompanied by adequate human supervision and verification of The information generated does not, in itself, constitute any infraction. The illegality arises when... The tool is used irresponsibly, without due professional diligence. resulting in an error that caused — or could cause — harm to the public interest.

7.3 Accountability and transparency

One question that remains open in both doctrine and practice is that of the obligation to disclose. In other words, whether the attorney has a duty to inform the court and the opposing party that they have used... AI systems in the drafting of legal documents. Some US courts already... They have issued rules to that effect, requiring parties to declare in a petition whether they have used AI. Generative in the elaboration process. In Brazil, the National Council of Justice initiated the debate on The topic exists, but there is still no binding rule imposing this duty on public legal counsel.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

From an ethical standpoint, there are relevant arguments both for and against obligation of disclosure. In favor, it is argued that transparency regarding the use of AI is essential. It allows the court and the opposing party to understand the process of drafting the document and to exercise greater discretion. scrutiny of its content. It is countered that the disclosure of the use of tools of No further work is required in relation to other technological tools, such as word processors. or legal databases, and that the attorney's responsibility for the document is complete. regardless of how it was developed.

The most balanced position, in our view, is that, in the absence of a rule that While imposing a duty, there is no obligation to disclose the use of AI as a support tool. However, when the use of AI results in an error identified by the prosecutor himself, ethics The professional mandates the immediate correction of the document, regardless of the consequences. have already been produced.

8 Ethical and Legal Limits of Delegating Activities Intellectuals to AI

8.1 The irreplaceability of professional judgment

An analysis of the potential benefits and risks of AI in State Advocacy leads to... This inevitably leads to the deeper philosophical question underlying the entire debate: do they exist? Intellectual activities inherent to the legal profession that cannot be delegated to systems. automated, for reasons that are not merely technical — linked to the current inability of the model—but, structurally, ethically and legally?

In our view, the answer is affirmative. The legal profession, as a legal practice, is... fundamentally an exercise in judgment — the ability to evaluate, in a context Specific and unrepeatable, what is the most appropriate course of action in light of the facts, the law, and the... values that govern the legal system. This judgment cannot be reduced to a process. pattern recognition, which is what LLMs do: it involves an understanding A genuine interpretation of legal norms, an ethical assessment of conflicting interests, and a responsibility for the consequences of the decision.

In the context of State Advocacy, this irreplaceable character of the judgment The professional aspect is even more pronounced. The prosecutor does not represent only the interests of the public.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

The immediate interest of the public entity in a dispute: it represents the primary public interest, which includes Observing the law, protecting public funds, and respecting citizens' rights. One No AI system, however sophisticated, does not possess—and cannot possess—that dimension of responsibility towards the collective interest.

8.2 Ethical Limits of Automation

From an ethical standpoint, delegating intellectual activities to AI raises questions. that legal deontology is still in the process of systematically addressing. The first clear limit is that of intellectual authorship: when signing a piece that you know you have created developed predominantly by an AI system without adequate supervision, the prosecutor This would implicitly attest to the authorship of a work that is not theirs — which could be considered a form of intellectual dishonesty incompatible with professional ethics. professional.

A second limit is that of the client's interest — in the case of State Advocacy, the Public interest. The use of AI without adequate supervision exposes the interest. represented by risks that the prosecutor would be able to avoid: the inclusion of arguments factually incorrect, the omission of relevant defenses, and the adoption of strategies procedural measures inappropriate to the specific case. The subordination of professional judgment to Automated exit from the model, without critical review, violates the duty of loyalty to the interest. represented.

A third limit is that of procedural loyalty. Article 5 of the Code of Civil Procedure. Article 77 establishes duties of good faith and procedural loyalty applicable to the parties and to... their lawyers. The presentation of AI-generated arguments that the prosecutor did not verified — and which may be factually incorrect — may constitute a violation of these duties, especially if it results in the presentation of non-existent precedents or Factual statements without documentary support.

8.3 Legal limits: what the current rule allows and prohibits

Under current Brazilian law, there is no rule that expressly prohibits... The use of AI in drafting legal documents. What exists are general principles that...

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

interpreted in light of the specific characteristics of the technology, they establish the limits of Use permitted.

These are clearly permitted uses: the use of AI as a research tool and... Support for drafting minutes, provided that the prosecutor provides human supervision. qualified and assume full responsibility for the content of the final piece; the use of AI for document analysis and synthesis; the creation of libraries of standardized prompts. to expedite the production of statements on repetitive matters.

These are uses that are clearly prohibited or ethically questionable: the submission of documents. Created by AI without human review; the insertion of confidential information or data. Personal data protected by the LGPD (Brazilian General Data Protection Law) in AI tools not approved by the regulatory body; use of jailbreak techniques to induce the model to produce content that the Ethical safeguards of the system would prohibit it; and attributing authorship to AI in signed pieces. by the prosecutor without the court's knowledge.

9 Guidelines for Safe Use by Attorneys

Summarizing the elements analyzed throughout this study, the following are proposed: The following are guidelines for the safe use of AI systems by state prosecutors:

1. Principle of Mandatory Human Supervision: every procedural document that includes Any involvement of AI systems in its development must be thoroughly reviewed. by the responsible attorney before signing. The review must be substantive — not merely formal — and includes the verification of all factual and legal elements. and relevant case law.
2. Principle of Source Verification: no jurisprudential, legislative or other references Doctrinal arguments generated by an AI system should be included in legal documents without verification. prior to the primary sources. The presumption, for purposes of institutional security, must be There's always the possibility that the model may have hallucinated the reference.
3. Principle of Proportionality of Use: the participation of AI in the creation of the piece must to be proportional to the complexity and impact of the case. In routine and complex cases Given the low cost, the use of AI-generated templates, with simplified review, may be suitable. In complex cases, with a large impact or with the potential to set precedents for others In some cases, the role of AI should be more limited and human oversight more rigorous.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

4. Principle of Confidentiality: Attorneys must strictly respect the
Institutional guidelines on what information can be entered into AI systems.
Classified information, personal data protected by the LGPD (Brazilian General Data Protection Law), and information about...
Sensitive procedural strategies should never be embedded in non-tools.
approved by the agency.
5. Principle of Full Responsibility: the attorney must be fully aware that he/she is
the sole person responsible for the content of the signed piece, regardless of the degree of
The participation of AI in its development. The use of AI does not mitigate or eliminate the
professional, disciplinary or civil liability.
6. Principle of Continuous Updating: given the rapid evolution of AI technology,
Prosecutors must stay up-to-date on the capabilities, limitations, and risks of
tools they use, participating in institutional training programs and
Seeking ongoing updates on the topic.
7. Incident Reporting Principle: any incident related to the improper use of
AI — identification of hallucination in previously submitted evidence, suspected prompt injection.
Information leaks must be reported immediately to the governance committee.
from the agency, so that corrective measures can be taken.

CONCLUSION

This article has demonstrated that generative artificial intelligence represents a
a significant and irreversible transformation in the work environment of State Attorneys,
with the potential to substantially increase the productivity of prosecutors in their activities.
drafting minutes, conducting case law research, and consolidating arguments. However,
This potential will only be realized beneficially if the technology is used with
awareness of its risks, within an institutional framework of rigorous governance and with
full preservation of the prosecutor's professional judgment.

The hallucinations of language models, the risk of prompt injection, the leak
The availability of sensitive information and the possibility of manipulating responses are real threats.
which require coordinated technical, procedural, and institutional responses. None
These threats alone are sufficient to justify prohibiting the use of AI in the legal profession.
of the State — such a stance would be as misguided as the uncritical and unrestricted adoption of

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

technology. The path forward is one of responsible regulation that maximizes the benefits and Minimize the risks.

Prompt engineering applied to public advocacy and the creation of libraries. Institutional standardized prompts represent concrete instruments for raising the quality of outputs generated by AI, while contributing to the Standardization of demonstrations and institutional protection against identified risks. Governance of AI use — with formal policy, specific committee, program of Training and periodic auditing are necessary conditions for the technology to be... incorporated responsibly into the daily operations of state legal aid agencies.

In terms of responsibility, the analysis undertaken allows us to conclude that the The attorney is fully responsible for the content of every document he/she signs. regardless of the participation of AI systems in their creation. The violation of duties of diligence and supervision, when resulting in harm to the public interest, may to establish disciplinary, functional, and civil liability. The recognition of this Responsibility should not inhibit the use of technology — on the contrary, it should encourage its use. thorough and technically sound.

Finally, it should be noted that the regulation of AI use in public advocacy is... This matter transcends the limits of each agency and demands coordinated action at the executive level. Federal and state. The issuance of guidelines by the Federal Council of the OAB, by the Council National Justice Department and the Federal Government itself — based on technical studies. In-depth analysis and comparative experience — this is an urgent measure for Brazil Develop a regulatory framework appropriate to the challenge posed by artificial intelligence. for the rule of law.

REFERENCES

BRAZIL. Constitution of the Federative Republic of Brazil of 1988. Brasília, DF: Federal Senate, Available at https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Accessed on: June 1, 2026.

BRAZIL. Law No. 8,112, of December 11, 1990. Provides for the legal regime of federal public servants. civilians Available at https://www.planalto.gov.br/ccivil_03/leis/L8112cons.htm. Accessed on: June 1, 2026.

BRAZIL. Law No. 8,906, of July 4, 1994. Provides for the Statute of the Legal Profession and the Brazilian Bar Association (OAB). the Available at https://www.planalto.gov.br/ccivil_03/leis/l8906.htm. Accessed on: June 1, 2026.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

BRAZIL. Law No. 13.105, of March 16, 2015. Code of Civil Procedure. Available at: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm. Accessed on: June 1, 2026.

BRAZIL. Law No. 13.709, of August 14, 2018. General Law on the Protection of Personal Data (LGPD). Available at https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Accessed on: June 1, 2026.

BRAZIL. National Council of Justice. CNJ Resolution No. 332, of August 21, 2020. This document addresses ethics, transparency, and governance in the production and use of Artificial Intelligence in the Judiciary and provides other measures. Available at: https://atos.cnj.jus.br/atos/detalhar/3429. Accessed on: June 1, 2026.

BRAZIL. National Council of Justice. CNJ Resolution No. 433 of 2023. Establishes guidelines for the use of artificial intelligence by the courts. Available at: https://atos.cnj.jus.br/. Accessed on: June 1, 2026.

BRAZIL. Superior Electoral Court. Electoral Judicial Investigation Action No. 0600814-85.2022.6.00.0000. Inspector General of Electoral Justice: Minister Benedito Gonçalves. Decision of April 14, 2023. Available at: https://www.conjur.com.br/2023-abr-18/tse-multa-advogado-peticao-baseada-conversa-chatgpt/. Accessed on: June 4, 2026.

BRAZIL. Court of Justice of Santa Catarina. 6th Civil Chamber. Fine resulting from False case law generated by artificial intelligence in a repossession appeal. Decision of February 19, 2025. Available at: https://www.tjsc.jus.br/web/imprensa/-/tjsc-multa-autor-de-recurso-por-jurisprudencia-falsa-gerada-por-ia. Accessed on: June 4, 2026.

BRAZIL. Court of Justice of Santa Catarina. 5th Criminal Chamber. Warning to lawyer for habeas corpus drafted by AI, with non-existent jurisprudence. Decision of February 2025. In: MIGALHAS. TJ/SC warns lawyer for HC drafted by AI with false jurisprudence. 2025. They are Paul, February 10th Available at: https://www.migalhas.com.br/quentes/424313/tj-sc-adverte-advogado-por-hc-feito-por-ia-com-jurisprudencia-falsa. Accessed on: June 4, 2026.

BRAZIL. Regional Labor Court of the 7th Region. 3rd Panel. Case No. 0000702-38.2024.5.07.0016. Fine for bad faith litigation and notification to the OAB-CE (Brazilian Bar Association of Ceará) regarding fictitious jurisprudence with evidence of AI generation. Decision of June 2025. In: MIGALHAS. TRT-7: suspicion of jurisprudence generated by AI and fine for the lawyer. São Paulo, June 12, 2025. Available at https://www.migalhas.com.br/quentes/432475/trt-7-suspeita-de-jurisprudencia-gerada-por-ia-e-multa-advogado. Accessed on: June 4, 2026.

BRAZIL. Regional Labor Court of the 12th Region. Labor Court of Concórdia. Labor lawsuit with initial petition containing non-existent decisions, doctrinal citations, and judge, generated by AI. Judge Daniel Carvalho Martins. Decision from October 2025. Available at https://portal.trt12.jus.br/noticias/autora-de-acao-e-multada-apos-advogada-inventar-jurisprudencia-e-desembargador. Accessed on: June 4, 2026.

BRAZIL. Regional Labor Court of the 2nd Region. Case No. 1001128-84.2024.5.02.0044. Fine for bad faith and notification to the OAB-SP (São Paulo Bar Association) for the use of fictitious jurisprudence generated by AI. Rapporteur: Judge Fernando César Teixeira França. Decision of February 2026. In: CONJUR. The use of false jurisprudence generated by AI results in a fine and notification to the OAB. São Paulo, February 16, 2026. Available at: https://www.conjur.com.br/2026-Feb-16/uso-de-jurisprudencia-criada-por-ia-gera-multa-por-ma-fe-e-oficio-a-oab/. Accessed on: June 4, 2026.

BRAZIL. Regional Labor Court of the 18th Region. Labor Lawsuit – Summary Procedure No. 0001204-10.2025.5.18.0121. Fine for bad faith and official letter to the OAB-GO (Brazilian Bar Association of Goiás) for the use of..

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

Regarding false case law attributed to AI. Decision from May 2026. In: ROTA JURÍDICA. Use of case law attributed to AI generates fine for bad faith and notification to the OAB of Goiás. 2026.
Available at: https://www.rotajuridica.com.br/uso-de-jurisprudencia-atribuida-a-ia-gera-multa-por-ma-fe-e-oficio-a-oab-de-goias/. Accessed on: June 4, 2026.

BRAZIL. Superior Court of Justice. Official statement: attempts to use *prompt injection.*
The investigations will take place at the Superior Court of Justice (STJ). Brasília: STJ, May 20, 2026.
Available at: https://www.stj.jus.br/sites/portalp/paginas/comunicacao/noticias/2026/20052026-attempts-to-use-prompt-injection-in-the-STJ-will-be-investigated.aspx. Accessed on: June 4, 2026.

BRAZIL. Regional Labor Court of the 8th Region. 3rd Labor Court of Parauapebas.
Judgment. Case No. 0001062-55.2025.5.08.0130. Substitute Judge: Luiz Carlos de Araújo Santos Júnior.
Decision of May 12, 2026. In: MIGALHAS. *Prompt injection*
Hidden in the initial petition: the case of Parauapebas. São Paulo, May 2026. Available at: https://www.migalhas.com.br/depeso/455925/prompt-injection-oculto-em-peticao-inicial-o-caso-de-parauapebas.
Accessed on: June 4, 2026.

BRAZIL. Regional Labor Court of the 8th Region. 3rd Labor Court of Parauapebas.
Judgment. Case No. 0001062-55.2025.5.08.0130. In: MIGALHAS. *Prompt injection* in the initial petition: the case of Parauapebas — Migalhas de Responsabilidade Civil.
São Paulo, May 2026. Available at: https://www.migalhas.com.br/coluna/migalhas-de-responsabilidade-civil/456362/prompt-injection-em-peticao-inicial-o-caso-de-parauapebas. Accessed on: June 4, 2026.

BRAZIL. National Council of Justice. Resolution No. 615, of 2025. Provides for the use of generative artificial intelligence in the Judiciary. Brasília: CNJ, 2025. Available at: https://atos.cnj.jus.br/atos/detalhar/6001. Accessed on: June 4, 2026.

BROWN, Tom et al. Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems, v. 33, p. 1877-1901, 2020. Available at: https://arxiv.org/abs/2005.14165. Accessed on: 1 June. 2026.

CHUI, Michael et al. The economic potential of generative AI: The next productivity frontier. McKinsey Global Institute, Available 2023. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier. Accessed on: 1 June. 2026.

CONJUR. Attempts to use *prompt injection* at the STJ will be investigated. São Paulo, May 20, 2026.
Available at: https://www.conjur.com.br/2026-mai-20/tentativas-de-uso-de-prompt-injection-no-stj-serao-investigadas/. Accessed on: June 4, 2026.

CONSTITUTIONAL AI PRINCIPLES. AI safety and alignment: key frameworks. In: ANTHROPIC Research Blog. San Francisco: Anthropic, 2023. Available at: https://www.anthropic.com/research. Accessed on: June 1, 2026.

DAFOE, Allan. AI Governance: A Research Agenda. Future of Humanity Institute, University of Oxford, 2018. Available at: https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf. Accessed on: June 1, 2026.

DATA PRIVACY BRAZIL. *Prompt injection* in the Judiciary exposes the cost of AI without governance.
São Paulo, June 2026. Available at: https://www.dataprivacybr.org/prompt-injection-no-judiciario-expoe-o-custo-da-ia-sem-governanca/. Accessed on: June 4, 2026.

HALLUCINATION LEADERBOARD. https://github.com/vectara/hallucination-leaderboard. Available in: Vectara.
Accessed on: June 1, 2026.

JI, Ziwei et al. Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, vol. 55, no. 12, p. 1-38, 2023. Available at: https://arxiv.org/abs/2202.03629.
Accessed on: June 1, 2026.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

JUSDOCS. *Prompt injection* in the process: fine, OAB and STJ inquiry. 2026. Available at: https://jusdocs.com/blog/prompt-injection-processo-multa-oab-stj-inquerito-ia. Accessed on: June 4, 2026.

KASPERSKY. What is *prompt injection?* How AI can be manipulated and what you can do. 2026. Available at: https://www.kaspersky.com.br/resource-center/threats/prompt-injection. Accessed on: June 4, 2026.

KATZ, Daniel Martin; BOMMARITO, Michael; BLACKMAN, Josh. GPT-4 Passes the Bar Exam. Social Science Research Network, Available at: 2023. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389233. Accessed June 1, 2026.

MIGALHAS. *Prompt injection* is a "police matter," warns Minister Salomão. São Paulo, June. 2026. Available at: https://www.migalhas.com.br/quentes/457223/prompt-injection-e-caso-de-policia-alerta-ministro-salomao. Accessed on: June 4, 2026.

MIGALHAS. *Prompt injection:* STJ investigates attempted AI manipulation in petitions. São Paulo, May. Available at https://www.migalhas.com.br/quentes/456424/prompt-injection-stj-apura-tentativa-de-manipulacao-de-ia-em-peticoes. Accessed on: June 4, 2026.

MIGALHAS. *Prompt injection:* AI created the vector, not the misconduct. São Paulo, May 2026. Available at: https://www.migalhas.com.br/amp/depeso/456010/prompt-injection-a-ia-criou-o-vetor-nao-a-ma-conduta. Accessed on: June 4, 2026.

MCNAUGHTON, David. Prompt Injection Attacks Against LLM-Integrated Applications. arxiv, 2023. Available at: https://arxiv.org/abs/2302.12173. Accessed on: June 1, 2026.

OPENAI. GPT-4 Technical Report. arxiv, 2023. Available at: https://arxiv.org/abs/2303.08774. Accessed on: June 1, 2026.

Brazilian Bar Association. Code of Ethics and Discipline of the OAB. Approved by Resolution No. 02/2015. Brasília: OAB, 2015. Available at: https://www.oab.org.br/content/pdf/legislacaooab/codigodeetica.pdf. Accessed on: June 1, 2026.

Brazilian Bar Association. Federal Council. Recommendation No. 1, 2024. Provides for the use of generative artificial intelligence in legal practice. Brasília: OAB, 2024.

OWASP — Open Worldwide Application Security Project. *OWASP Top 10 for Large Language Model Applications 2025.* Available at: https://owasp.org/www-project-top-10-for-large-language-model-applications/. Accessed on: June 4, 2026.

PEREZ, Ethan; RIBEIRO, Marco Tulio. Ignore Previous Prompt: Attack Techniques for Language Models. arxiv, 2022. Available at: https://arxiv.org/abs/2211.09527. Accessed on: June 1, 2026.

ROMM, Tony. A lawyer cited ChatGPT in court. The judge was not amused. The Washington Post, May 25, 2023. Available in: https://www.washingtonpost.com/technology/2023/05/27/chatgpt-ai-lawsuit-avianca-airlines/. Accessed on: 1 June. 2026.

VASWANI, Ashish et al. Attention Is All You Need. Advances in Neural Information Processing Systems, v. 30, 2017. Available at: https://arxiv.org/abs/1706.03762. Accessed on: 1 June. 2026.

WEI, Jason et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. Advances in Neural Information Processing Systems, v. 35, 2022. Available at: https://arxiv.org/abs/2201.11903. Accessed on: 1 June. 2026.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google

Year VII, v.1 2026 | Submission: 04/06/2026 | Accepted: 07/06/2026 | Publication: 10/06/2026

WHITE, Jules et al. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. arxiv, 2023. Available at: https://arxiv.org/abs/2302.11382. Accessed on: 1 June. 2026.

WORLD ECONOMIC FORUM. Applying AI in Legal Contexts: Opportunities, Risks and Governance. Geneva: WEF, 2024. Available at: https://www.weforum.org/reports/. Accessed on: June 1, 2026.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.