.**UBA**derecho   **IA**LAB
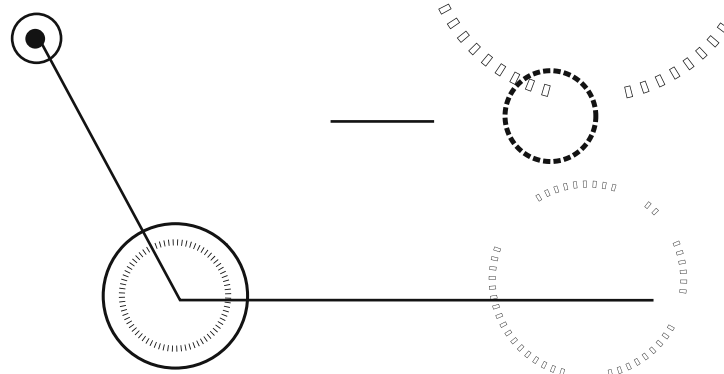
Guidelines for the use of ChatGPT and text generative AI in Justice

Guidelines for the use of ChatGPT and text generative AI in Justice

General Management: Juan Gustavo Corvalán and Mariana Sánchez Caparrós

UBA-IALAB research team: Giselle Helleg, Carina Papini, Melisa Raban and Antonella Stringhini.

Dyntec-Lab research team: Marco Rossi, Franco Orellana

I. Executive Summary

The emergence and general public availability of disruptive models such as ChatGPT, GPT-4, Bard, DALLE-2, Midjourney and Stable Difussion, has given rise to a new era of creation and manipulation of text, image and videos. This is an evolutionary leap in artificial intelligence, based on the generation of synthetic digital content.[1]

Generative AI is a subfield of research and development within the artificial intelligence ecosystem. It is evolving rapidly and focuses on generating images, music, text, videos, voices and computer code, from natural language text input provided by the user (*prompt* or instruction).

The new AI assistants also make it possible to make multiple legal and judicial tasks more efficient, while at the same time they can enhance various aspects linked to the underlying legal reasoning.

From the daily management of simple procedural acts, to the possible generation of documents (e.g. certificates, official letters, orders, notes, dispatches and resolutions, among others). Also, these large language models (LLM) such as ChatGPT, can address other more sophisticated tasks that carry multiple challenges and associated risks.

The massification and making generative AI available for free for a vast public, has to be qualified by the sensitivity of the state function performed by those who make up the Judiciary. Various constitutional principles and rights come into play here, while the principle of competition could be tense or violated.

The approach that we propose in this guide is based on the following five postulates:

---

[1]The term synthetic digital content refers to artificially generated information that can be used in place of real historical data to train AI models. See Korolov, Maria, "What is synthetic data? Data generated to help your AI strategy", March 15, 2022, available at: https://www.ciospain.es/big-data/que-son-los-datos-sinteticos-datos-generados-para-ayudar-a-tu-estrategia-de-ia

Generative AI has the ability to alter the data on which they based their learning, at the same time that they can generate others that never existed. This is a type of recombination that generates novel content from the generation or addition of synthetic data. See Corvalán, Juan G.,*"Generative artificial intelligence like ChatGPT: A new renaissance?* The Law, June 5, 2023.

1) Literacy, awareness and permanent training based on the evolution of the state of the art of generative AI.

2) Responsible use, based on mitigating risks associated with large generative AI language models.

3) Human in the Loop: Essential human control before and after. Prohibition of delegation of decision making.

4) Understanding the intrinsic and external limitations of LLMs such as ChatGPT (inventions, hallucinations, inconsistencies, negative biases, among others).

5) Strategic impact assessment according to the scope of application.

In recent times, situations of inappropriate use have been made known, such as the case of the lawyers in the "Mata vs. Avianca Airlines., Inc.", which went viral because the representatives of one of the parties had cited a series of non-existent judicial precedents to support a claim, resulting in them being required by Justice to give explanations in this regard.

On that occasion, the aforementioned lawyers admitted to the court that it was an error due to the misuse of ChatGPT by one of them, a professional with a long career, who relied entirely and trusted the content generated by the system, without subsequently verifying that result.[2]

These types of situations have paved the way for other judges, such as the case of a federal judge in the Northern District of Texas, to establish, as a rule, the need for the presentation of an explicit and specific statement by the lawyers, in which they indicate that no part of their presentation was written by an AI or, if so, inform that the content has been verified by a human being as to its accuracy and veracity of sources, prior to being presented to the court.[3]

This example shows that the incorporation of text generative AI as support for the execution of daily legal or judicial tasks brings with it a great challenge for

---

[2] Weiss, Benjamin,*"This is what happens when your lawyer uses ChatGPT," Infobae*, May 28, 2023, available in: https://www.infobae.com/thenewyorktimes/2023/05/28/esto-es-lo-que-ocurre-cuando-tu-abogado-utiliza-chatgpt/ [accessed 6/11/2023].

[3] Jorge, Miguel, "A judge orders that all content generated by an AI in court be declared", May 31, 2023, inhttps://www.msn.com/es-us/noticias/other/un-juez-ordena-que-todo-el-contenido-generado-por-una-ia-en-el-tribunal-se-declare/ar-AA1bVraM [accessed 6/11/2023].

organizations. Every new technology allows us to discover new kinds of uses while entailing new responsibilities.[4] Generative AI is no exception.

Based on the work we do at UBA IALAB on the impact of ChatGPT in different areas, including the legal field, we believe that it is a critical moment to carry out a "controlled and strictly supervised experimentation", ushered by guidelines that are updated and modified over time with constant feedback from different stakeholders. The MIT Working Group is also working with this same line of thought in regards to the use of generative AI in the legal field.[5]

Below we take the first step to establish a general guide, in the form of guidelines and recommendations, as a starting point in relation to the responsible use of this disruptive technology.[6]

II. Goal

We propose a series of guidelines and recommendations for the responsible, ethical, appropriate and diligent use of text generative AI in the justice service.

This type of technology is what is behind the systems that are available to the general public through different applications such as ChatGPT, GPT4, Bing and Bard, among others, some of which are free to access.

Specifically, it is about making known the inherent limits and possibilities of these intelligent systems, and then providing guidelines so that users in the justice sector can use them in line with the duties and values that must be protected in the exercise of their function. Here you have to balance the possible benefits with the risks and potential harms.

---

[4] Center of Humane Technology, "The A.I. Dilemma", March 9, 2023, available at https://www.youtube.com/watch?v=xoVJKj8lcNQ [accessed 6/7/2023].

[5] Task Force on Responsible Use of Generative AI for Law, available at: https://law.mit.edu/ai [accessed 6/11/2023].

[6] In line with this, the research that gave rise to the book was promoted from the Innovation and Artificial Intelligence Laboratory of the Faculty of Law of the University of Buenos Aires.*ChatGPT vs. GPT-4: imperfect by design?*, UBA IALAB - Thomson Reuters La Ley, March 2023, available at:https://ialab.com.ar/wp-content/uploads/2023/03/Libro-ChatGPT-vs-GPT-4.-UBA-Thomson-Reuters-La-Ley.pdf There, a specific investigation was carried out on two generative text AIs to analyze their inherent possibilities and limitations. Through various tests, an attempt was made to obtain information to guide the reader towards a more appropriate use of these systems.

For the rest, this first version of the guide is developed under an open and iterative logic, in order to make it known and invite a broader and more diverse set of interested parties to participate in its eventual improvement and completeness, at the same time than to keep it permanently updated (See point X).

III. Conceptual framework. Possibilities and limits of large language models

LLMs emerged strongly towards the end of 2022 with the launch and general public availability of ChatGPT, and then GPT-4 in March 2023.

These two models, from the company OpenAI, paved the way into a paradigm shift. On one hand, the provision to the user of a tool with the potential to transform various industries and sectors; on the other, democratization for access. Anyone with a device and an internet connection can now use them freely, regardless of whether or not they understand the underlying technology.

LLMs, such as ChatGPT and GPT-4, are AI systems that can process natural language in the form of a conversation. Under this logic, they allow the user to generate textual content in various formats (e.g. poetry, news, questions, summaries, etc.) from an instruction or requirement provided to the system in natural language, commonly known as *prompt*.

From a more technical perspective, they are generative AI systems[7] trained on text string prediction tasks. Meaning they are trained to predict the probability of occurrence of a *token* (which can be a character, a word, or a string of words) given its preceding or surrounding context.[8]

---

[7] Generative AI is a subfield of AI research that focuses on training algorithms that can be used to generate synthetic (artificial) content that resembles that created by humans, such as text, images, graphics, computer code, etc. Generative models can analyze the patterns and structures present in the training data and learn from them to produce new content that shares characteristics with that original data, allowing the content to appear authentic and similar to what a person would produce. Expand in Fezari Mohamed, Al-Dahoud Ali, Al-Dahoud Ahmed, "Augmenting Reality The Power of Generative AI", preprint, May 2023, in https://www.researchgate.net/publication/370821758_Augmanting_Reality_The_Power_of_Generative_AI" [accessed 6/22/2023].

[8] Bender Emily, McMillan-Major Angelina, Gebru Timnit, Shmitchell Shmarga, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, New York, March 2021, https://doi.org/10.1145/3442188.3445922.

In the case of ChatGPT, it is a generative model based on the Pre-Trained Generative Transformers (GPT) architecture, which can process sequences of elements (such as the words in a sentence) using a *deep learning* architecture that makes it easier for you to pay attention to the different parts of a sequence of words while processing it, achieving greater efficiency and precision in your tasks.[9]

ChatGPT has been fine-tuned using Human Feedback Reinforcement Learning (HRFL) technique, whereby human annotators teach the model to be more accurate by rewarding and penalizing its results. This circumstance means that, from its training, the model is subject to the subjectivities of our species.[10]

As recognized by Sam Altman, CEO of OpenAI, and as we have verified from the IALAB in the research carried out this year[11], LLMs in general, and ChatGPT and GPT-4 in particular, are often imperfect and limited by design. That is, they are deployed to the world with knowledge of their defects so that society adapts to the product.

In fact, within the limits of these systems, it is noted that:

a. Sometimes they give coherent and convincing answers, which imitate confident style and expert jargon, but are incorrect or false (they hallucinate).
b. In their responses they sometimes reflect prejudices, stereotypes, beliefs and negative social values present in their training data (bias).
c. Sometimes they are not solid and fail relatively frequently when given tasks that involve logical reasoning.
d. These are systems that are very sensitive to adjustments in the formulation of input phrases or sentences.

---

[9] Expand in *ChatGPT vs. GPT-4: imperfect by design?*, UBA IALAB - Thomson Reuters La Ley, March 2023, available at:https://ialab.com.ar/wp-content/uploads/2023/03/Libro-ChatGPT-vs-GPT-4.-UBA-Thomson-Reuters-La-Ley.pdf

[10] Expand in *ChatGPT vs. GPT-4: imperfect by design?*, UBA IALAB - Thomson Reuters La Ley, March 2023, available at:https://ialab.com.ar/wp-content/uploads/2023/03/Libro-ChatGPT-vs-GPT-4.-UBA-Thomson-Reuters-La-Ley.pdf, p. 32.

[11] Expand in *ChatGPT vs. GPT-4: imperfect by design?*, UBA IALAB - Thomson Reuters La Ley, March 2023, available at:https://ialab.com.ar/wp-content/uploads/2023/03/Libro-ChatGPT-vs-GPT-4.-UBA-Thomson-Reuters-La-Ley.pdf

e. Some of these systems have knowledge limited to a certain date (e.g. ChatGPT).

f. They are excessively detailed and over-explain.

g. They make assumptions about facts.

Arthur C. Clarke many decades ago stated: "Any sufficiently advanced technology is indistinguishable from magic." In the judicial branch, this statement is not sustainable from any point of view when evaluating the usefulness of a tool or technology. In fact, they are not artificial oracles who know everything and can answer everything well.

Also, another fundamental starting point to consider is that these LLMs deepen the notion of "black boxes" that we have worked on in multiple investigations, projects and conferences. This trait has two relevant dimensions to consider. On one hand, it is not known with what data the system was trained. There are various hypotheses about the sources used, but there is no reliable information that specifically indicates which databases they use to generate the answers.

In this dimension, we also do not know what parameters the company applies to the language models. That is, the rules that adjust the outputs of a system. An obvious example might be this: if I ask it to tell me how to be Hitler, the system refuses to answer. This is not "natural" in the system but is a specific configuration made by the programmers.

In another dimension, the black box responds to the way in which data is processed. It is a subspecies of artificial neural networks that mathematicise knowledge based on probabilistic criteria. While they process the data, they do not know the specific step by step of how they arrive at a certain result. We have explained this phenomenon in various articles and research-.[12]

This ultimate characteristic makes it difficult and, in certain cases, eliminates the possibility of explaining with adequate detail the reason for the decision, in order to justify and properly motivate decisions in cases where this is constitutionally or

---

[12] See Ienca, Marcelo and Vayena, Effy, "AI Ethics Guidelines: European and Global Perspectives", in CAHAI Secretariat (Comp.), Towards Regulation of AI Systems, December 2020, p. 51, available at: https://www.coe.int/en/web/artificial-intelligence/-/- toward-regulation-of-ai-systems- [accessed on 5/6/2023]. Also, CORVALAN, Juan- "Generative artificial intelligence like ChatGPT: A new Renaissance? An explosion of collective human intelligence on the shoulders of AI" - Published in: LA LEY 06/05/2023, 1 Online quote: TR LALEY AR/DOC/1278/2023

legally required. Ultimately, this prohibits any alternative to decisional delegation and, in some way, limits the role of this type of tools to that of an assistant.

IV. Possible uses of text generative AI in Justice

The use of text generative AI in Justice can be presented as a genre that includes two species:

1. Transversal uses: they can occur in relation to any procedural stage and instance and in relation to any writing;
2. Specific uses: they are presented for specific tasks or documents that must be planned and dictated in the different procedural stages, instances and jurisdictions (specialty criteria of the subject involved).

Below we will provide a list of possible transversal uses of AI in Justice, as well as specific uses. This list, although broad, is merely illustrative, given that this subfield of generative AI is dynamic and is constantly developing and evolving.

It is key to promote continuous improvement to reveal other possible uses, just as it is important to consider those that should be discarded, and it is also critical to make constant reviews about precision, coherence, hallucinations and all issues linked to the limits of large models. of generative AI language.

1. What would be some possible transversal uses of generative AI in Justice?

Transversal uses are those that can occur independently of the procedural stage and instance that a judicial case goes through and in relation to any writing. It includes the pre-judicial stage as that that occurs when the process is ruled.

Among the possible transversal uses identified, so far, we have found the following:

a. Search for existing information (National Constitution, treaties, doctrine, jurisprudence, etc.).
b. Search within texts (National Constitution, treaties, doctrine, laws, jurisprudence, etc.).
c. Request for ideas or alternatives to problems or judicial conflicts.

d. Synthesis of different types of judicial documents (e.g. lawsuits, pleadings, sentences, doctrine, etc.).

e. Interpretation, evaluation or weighing of legal rules or principles applicable to a specific case.

f. Generation of models or templates for resolutions (e.g. simple rulings, opinions, etc.).

g. Make analogies or metaphors about arguments, claims or possible explanations linked to the content of a judicial decision.

h. Combination of legal information with information from other disciplines (e.g. art, literature, cinema, etc.).

i. Conceptual distinctions and combination of legal arguments.

j. Strengthening and deepening of legal arguments that are presented as a starting point for the system.

k. Analysis of judicial or legal documents (e.g. contracts, writings, demands, resolutions, etc.).

l. Compare data or information between resolutions, legal regimes or other documents.

m. Assessment, interpretation and/or consideration of the origin of certain claims.

n. Respond to emails, notes and letters.

o. Translation of documents.

p. Improve writing, apply clear language/inclusive language/synthesize/remove gerunds.

2. **Which would be some possible specific uses to generative AI in Justice?**

Specific uses are those that are presented for specific tasks or writings that must be planned and dictated in the different procedural stages, instances and jurisdictions (specialty criteria of the subject involved).

Among the possible specific uses identified, so far, we have found the following:

a. Redaction of drafts of simple orders and sentences.

b. Relationship between legal texts provided by the user (e.g. demand and response to the demand).

c. Relationship between judicial texts and texts provided by the user (e.g. sentence and appeal).

d. Identification of the claims and requests made in writings (e.g. in a complaint or in a response).

e. Enumeration of the means of proof proposed in writings.

f. Suggestion of new means of proof or expansion of means of proof.

g. Summary of the object of the claim.

h. Suggestion for improvements in the writing and content of judicial texts.

i. Analysis and evaluation of evidence produced in relation to the facts under debate and claims of the parties.

V. Relevant considerations regarding the potential uses of text generative AI identified in this guide

It is important to keep in mind that the identification of possible transversal and specific uses carried out in this guide does not necessarily mean that its implementation will yield correct, precise, safe, useful and/or adequate results.

Indeed, the current state of evolution of text generative AI, initially addressed in the book "ChatGPT vs. GPT-4: imperfect by design?", shows that we still face systems with important limitations that many times, to achieve good performance, require a user with good *prompting* skills.

That is, with knowledge and skills that allow them, through the instructions they provide to the system, to guide it in carrying out the task of processing and generating natural language, with or without legal content, to obtain more efficient, useful, effective and accurate results.

Notwithstanding this, it should not be lost of sight that sometimes the limits go far beyond the capacity of the human person to interact with the system.

VI. Specific guidelines and recommendations for the use of AI in Justice

VI.1. Introduction

The responsible, ethical, appropriate and diligent use of text generative AI by judicial officers demands a holistic approach that integrates specialized knowledge;

transparency; human supervision; regulatory compliance; protection of data and information of the organization and third parties; addressing challenges related to biases and maintaining constant and ongoing critical and ethical judgment on the part of the system user.

In essence, a balanced approach must be adopted between the adoption of generative AI, which is presented as a work tool with the capacity to enhance, amplify and complement human intelligence, and the preservation of the fundamental duties, values and principles that go through the judicial function.

Firstly, these guidelines have as their conceptual framework the map of ethical documents that has been synthesized and that form part of the book titled "Tratado de Inteligencia Artificial y Derecho", published by Thomson Reuters-La ley, in November of 2023.[13]

From this perspective, we propose an illustrative list of guidelines that every judicial agent, regardless of their position or hierarchy, should comply with to make responsible, ethical, appropriate and diligent use of text generative AI when using it as a tool at work.

At the same time, others that are stated aim to promote that people who use generative text AI in Justice can make the most of the capacity of these systems and obtain more accurate, useful, precise and satisfactory responses and, at the same time, optimize the interaction with the AI system to make it more efficient (fewer interactions/better results).

The latter ones will help agents make ChatGPT or other AI that are used, a useful, agile and versatile assistant for the development of some of your daily tasks that involve writing text in different formats, analyzing and solving problems and synthesizing documents, among others.

In all cases, the proposed guidelines will allow agents to make ChatGPT or other AI that are used, a useful assistant for the development of some of your daily tasks that involve writing text in different formats, analyzing and solving problems, and synthesizing documents, among others.

---

[13] See Corvalán, Juan G. (Dir.), *Tratado de Inteligencia Artificial y Derecho*, Buenos Aires, La Ley, 2023, available at: https://www.thomsonreuters.com.ar/es/soluciones-legales/tratado-de-inteligencia-artificial-corvalan.html.

Below, the identified guidelines are specified and grouped. Also, we provide, as a suggestion, some specific recommendations that can be adopted by organizations to make them effective in their daily practice.

VI.2. List of guidelines and recommendations

i. In relation to the data of the organization and third parties

i.1. When using documents or information of the organization as input: protect and ensure confidentiality

Court documents often contain sensitive, confidential or private information, so we must ensure that we implement appropriate security measures to protect the privacy and confidentiality of the organizations and third-party data contained in those documents while using generative AI.

The use of text-generative artificial intelligence in the judicial field should not lead to the processing by third parties of personal data, nor organizational data, so it is important to comply with data protection regulations, but also with the rules of reserve of the judicial power itself to guarantee the confidentiality of the information.

Implementation Tips:

- Design, together with specialists in data protection and information security, a scheme of best practices in risk management to guarantee the confidentiality of the information of the organization and third parties contained in documents and systems.
- Best practices should include techniques for users to learn how to implement data anonymization mechanisms to ensure the protection of the organization and third-party information.
- Design awareness campaigns and ongoing training for users on the security and privacy risks associated with text-generative artificial intelligence used in the judicial field.
- Establish strong confidentiality agreements with agents to safeguard the data they manage.

ii. In relation to the people of the organization

ii.1. Analyze the level of knowledge of users and design awareness and knowledge plans for generative AI

The awareness and knowledge of text generative AI among the people who are part of the organization are essential to harness its potential; ensure appropriate, ethical, legal and successful use of this tool, and avoid overconfidence.

It is important that the organization's authorities know the degree of knowledge that the agents have about the operation, possibilities and limits of text or image generative AI.

A holistic view of the level of knowledge of the organization will allow you to identify training needs and appropriately address existing knowledge and skill gaps.

This can be done through the use of various strategies and means of communication, such as campaigns, talks and training that focus on explaining the three verticals referred to (functioning, possibilities and limits), as well as the human roles and responsibilities in the process, the risk prevention and best practices.

Implementation Tips:

- Design and implement specific communication campaigns on text generative AI in the Judiciary.
- Implement training and continuing education programs for different levels of users (initial, intermediate, advanced) focused on understanding technology and its ethical implications.
- Provide a clear description of the guidelines and recommendations for the use of text generative AI in the Judiciary.
- Train on risk management policies and practices adapted to text generative AI.
- Emphasize the role of people in validating and verifying AI-generated texts.
- Establish multidisciplinary consultation teams so that they can address the study and work of complex use cases.

iii. About the results of the AI system

iii.1. Iterate responses with appropriate human supervision focused on manually reviewing and editing responses

Human supervision is essential to ensure the accuracy, quality and ethics of AI-generated responses, correct and improve them, prevent bias, inappropriate, incorrect or illegitimate content.

Therefore, it is important to permanently emphasize that human beings will always continue to be control and decision-making agents. This will mitigate fears and increase confidence in the technology, while collaborating in its legal, ethical, responsible and appropriate use.

Systems such as ChatGPT, GPT-4 and Bard, among others, have been designed specifically for text generation tasks in conversational format. They are not optimized to be used as a search engine.

These are models with an adequate structure to generate coherent answers to questions and hold natural conversations, so they may present limitations if they are used to search and retrieve specific information with a precise and reliable result.

This is why the use of text generative AI, at least for the moment, can lead to the generation of coherent and convincing answers, which appropriately use technical language, but which may eventually be inaccurate, incomplete or based on outdated or false information.

Implementation Tips:

Ensure adequate human supervision to:

- Always use verification mechanisms and subsequent human validation of responses to guarantee their accuracy and reliability.
- Always use verification mechanisms and subsequent human validation of responses to control the sources cited when dealing with systems connected to the Internet.
- Correct errors, improve quality and ensure the appropriateness of responses.
- Check that they do not contain false information.

- Check that they do not contain inappropriate or offensive content.
- Ensure their ethical adequacy (e.g. checking that they do not reflect negative stereotypes or prejudices, etc.).
- Ensure its adaptation to current regulations and to the specific case according to the facts.
- Improve your precision.
- Approve methodologies for the preparation of *prompts* appropriate for each case.

iii.2. Consider AI-generated documents as drafts

In addition to the limits indicated, related to the risks that these systems present in terms of generating erroneous, biased or meaningless responses, it must be kept in mind that LLMs act as "black boxes," which brings serious difficulties. In principle, it is not possible to understand and justify how the AI arrived at a specific result or decision.

As a result, when they are used to generate judicial documents, they should always be used as an assistant or co-pilot to generate a draft that will speed up the task of final preparation of the document which, in all cases, must be left to a human person who will control its content by applying its *expertise*.

Implementation suggestions:

1. Define specific guidelines for the use of generative AI in the drafting of judicial documents (e.g. in which cases, with what *prompts*).
2. Work on the design of appropriate *prompts*, that provide the system with a context, provide terminology, contextualize situations and provide examples, among others, to improve the quality of the drafts generated.
3. All drafts generated by AI must be reviewed and edited by people qualified in the subject matter being worked on. This is to ensure their accuracy and consistency with applicable standards and rules.
4. Regularly monitor and evaluate the performance of generative AI used in the generation of judicial documents through user feedback in order to introduce improvements in the way it is used and identify new use cases.

iii.3. Evaluate and mitigate biases in the responses generated

Text generative AI systems can produce biased responses that reflect gender stereotypes, prejudices, beliefs, or negative social values,[14] so people who work in the Judiciary must be aware of this problem and have tools to evaluate and eliminate them from the results generated, in order to mitigate their negative impact, avoid discrimination or unfair differences in treatment and representation, and ensure fairness in the use of generative AI.

Implementation suggestions:

1. Train and sensitize agents of the Judicial branch so that they understand how generative text AI systems work and the problems they present in relation to bias.
2. Encourage collaboration between experts in AI ethics and judicial ethics to develop protocols for the evaluation and mitigation of bias in AI-generated responses.
3. At the individual level, carry out a rigorous evaluation of the responses based on the applicable ethical and regulatory standards to ensure they are impartial and do not incur in discrimination or unfair or unacceptable differences in treatment and representation of groups.
4. At the team level, ensure careful review of AI-generated responses, especially when used in the context of the exercise of judicial function, to corroborate the impartiality/objectivity of the generated response and correct any bias or inaccuracy identified.

iv. Regarding prompting

iv.1. For *prompting* in general

In the context of text generative AI, a *prompt* is an instruction or initial phrase that is provided to the LLM to generate text and achieve specific results such

---

[14] Expand on "ChatGPT vs. GPT-4: imperfect by design?", UBA IALAB - Thomson Reuters La Ley, March 2023, p.15-16, available at:https://ialab.com.ar/wp-content/uploads/2023/03/Libro-ChatGPT-vs-GPT-4.-UBA-Thomson-Reuters-La-Ley.pdf

as writing texts in different formats (e.g. news, emails, poems, etc.), automatic translation, dialogue generation, among many others.

In other words, it is a text input provided in natural language by the user of the system that serves to guide the latter in generating the desired response. For example: "Answered the email that I am going to provide you below in a gentle but formal tone, in a maximum of 10 lines. Informal to the person that I will not be able to attend the meeting on 7/1/2023 because I have other commitments."

The *prompt* is a way of communicating with text generative AI systems, so it is important to adjust and formulate them appropriately to obtain better results. Also, it is important to be aware that each model can respond differently to the same *prompt*, and that the same system can generate new results for the same prompt.

Specific suggestions:

- **Be clear, precise and concise:** provide clear, specific and, to the extent possible, concise instructions.

  Example: Unclear and unspecific *prompt*: "Write me 15 lines about AI and Law."/ Clear and specific *prompt*: "Write me 15 lines about AI and Law. In particular about the relationship that exists between both disciplines, what are the points of contact between both and how they influence each other."

- **Provide sufficient details:** provide the details so that the system can better process the instruction and more appropriately comply with it.

  Example: *Prompt* not detailed: "Write me 15 lines about music." / Detailed *prompt*: "Write me 15 lines about the influence of national rock on Argentine culture in the 1980s."

- **Set a context:** provide an adequate context so that the system can adjust to the topic or situation raised.

  Example: *Prompt* without context: "Write an email to invite to a meeting." / Contextualized *prompt*: "Write an email in a gentle and formal, brief tone, to

invite Juan Pérez's work team, to his office, to address different topics that have been worked on in the last month."

- **Indicate recipients:** indicate who are the intended/target users of the text to be generated.

  Example: *Prompt* without intended/target users: "Summarize this text using clear language." /*Prompt* with intended/target users: "Summarize this text using clear language. Aim it at 15-year-old teenagers." They could also be: non-professionals; kids; teenagers; older adults; etc.

- **Set a tone:** indicate the tone that the text to be produced should have. For example: formal, informal, funny, persuasive, assertive, etc.

- **Use examples and keywords:** include examples or key words to clarify and illustrate the objective sought with the instruction.

  Example: *Prompt* without examples: "Describe a scene of two children playing." /*Prompt* with examples: "Describe a scene of two children playing. For example, how the brothers from the series "Loud" play *House.*"

- **Experiment and iterate the *prompt*:** sometimes, minimal adjustments in the instruction or requests for clarification, precision and provision of new details to the generated result, through the use of new *prompts* or the iteration of the one you are using, lead to more precise and interesting results.

- **Indicate a role:** request the system to assume a specific role prior to giving the instruction.

  Example: *Prompt* without role: "Describe a scene of two children playing." /*Prompt* with role: "Assume the role of a science fiction writer for children. Describe a scene of two children playing. For example, how the brothers in the series play *Loud House.*"

- **Request a format:** indicate the desired output format for the text. For example: draft, writing, outline, comparative table, dialogue, etc.

- **Indicate an output writing style:** request the desired writing style for the text to be generated. For example: clear language, inclusive language, inclusion of gender perspective.

  Example: *Prompt* without asking for style: "Describe a scene of two children playing" / *Prompt* with style: "Describe a scene of two children playing, apply a gender perspective when considering the *game they play."*

- **Citation of sources:** for systems connected to the internet (e.g. Bing), request citation of all sources that support the content of the generated response and then check them.

iv.2. For *legal prompting*

A *legal prompt* is an instruction or initial phrase that is provided to the LLM for the generation of text related to legal matters.

The *legal prompt* can be especially useful for applying generative AI to tasks related to the legal field such as drafting contracts, analyzing facts, generating legal arguments and counterarguments, among others.

Suggestions specific to achieve a good *legal prompt:*[15]

- **Be specific and clear:** Provide precise and clear instructions on the legal topic you are going to work on.

- **Set the context:** Include relevant information to provide adequate context to the system, such as, for example, details of facts to be analyzed, non-relevant circumstances, applicable principles, legal references, among others.

- **Assign a tone:** Specify the desired output tone for the general text, appropriate to the task being performed. For example: formal, informal, assertive, persuasive, funny.

- **Request a format:** Request the desired output format for the text. For example: draft, writing, outline, comparative table, dialogue, etc.

---

[15] Some of the suggestions provided here have been contributed by Dr. Sebastión Chumbita in his work on legal *prompt engineering*, published on May 31, 2023. Expand in Chumbita, Sebastián, "Legal prompt *engineering*", *Law 2023-C*, May 31, 2023.

- **Indicate a role:** Request the system to assume a specific role ("act as") prior to giving the instruction. For example: role of judge, role of specialist in law and technology, etc.

- **Provide the desired objective:** Indicate to the system what goal or purpose is sought to be achieved with the response. For example: inform, reconcile, resolve, analyze, etc.

- **Include keywords:** Incorporate keywords into the instruction to direct the response generated towards the desired objective. For example: appeal, demand for support, arguments in a labor trial, oral arguments, etc.

- **Set limits:** Specify restrictions on the answer. For example: number of words or characters, line limit, do not use generic masculine or grammatical masculine.

- **Indicate the target audience:** Define who the target recipients of the text are to achieve personalized content. For example: "write a summons for an interview to make contact with a minor."

- **Request fonts:** In models connected to the Internet (e.g. Bing), request the citation of all sources that support the content of the generated response and then check them.

- **Request consideration of multiple perspectives and counterarguments:** Ask you to take on more than one role and consider multiple points of view, opinions, and counterarguments on the same topic. See example of *superprompt* technique.

Example of a *prompt* for *compliance*:

Hello ChatGPT, [insert text describing the facts of the case in sufficient detail].I need you to determine the legality of [insert the text that describes the facts to be analyzed in sufficient detail], based on [indicate the regulations under which the analysis is intended to be carried out].

It is likely that after the first response we must iterate with the AI to refine and/or adjust the result, as well as to clarify doubts or unresolved questions that arise from that initial response.

iv.3. *Legal prompting* for procedural acts

Text generative AI can also be used as an assistant for dictation of procedural acts that respond to requests made by the parties during the process.

This allows not only to simplify the way in which the response is submitted to an office, but also to reduce its length and also to create models that can incorporate numerous determining procedural variables.

To obtain accurate, controlled, coherent and reviewable results it is advisable to have a conversation which complies with the specific recommendations for a good *prompting legal*, applied to a conversation context using the methodology suggested below.

Specific suggestions to achieve a good *legal prompt* for procedural acts:

***Prompt* to generate simplified order models:**

Hello ChatGPT. Simplify the following text [office model], maintain a wording that allows it to be used as part of a model, use legal language, keep data that refers to telephone numbers, law numbers, deadlines, amounts of money, names of entities.

***Prompt* to generate office models with options:**

1. Because generative agents do not have specific procedural knowledge, and because they can respond to conceptual explanations, dispatch models can be established with indications of the procedural circumstances that would motivate its dictation. In this way, you can approach a conversation with the system that will allow you to arrive at a useful response in the course of the conversation.

A first *prompt* model with numerous determining procedural variables is provided below:

Hello ChatGPT. I am going to assign you two prompts to work with options later depending on the next step that the process must follow. Option 1: ["in case of compliance with the presentation of… it is resolved…"]. Option 2: ["in case of not complying with the presentation of… it is resolved."].

2. Taking as a starting point a *prompt* with numerous procedural variables that determine the dictation of one or another decision, the system can be asked to prepare a list of questions that constitute the procedural reasoning on which a decision must be made. The result is more precise if the number of options is indicated.

A second *prompt* model is provided below:

Based on the following text [prompt with numerous determining procedural variables], prepare a questionnaire of the questions that must be taken into account in each of the XX options.

3. If the determining procedural variables are appropriately written and the questions in the options questionnaire are properly answered, an automatic response from the AI can be achieved, consistent with the presentation of the party and the procedural moment being passed. To motivate a response of these characteristics, it is necessary to request it together with the answers to the previous questionnaire.

A third *prompt* model is provided below:

Now I need you to use the answers from the following list [responses to the procedural variables questionnaire] and complete the dispatch model [dispatch/providence model with determining procedural variables].

4. It is possible that the result of the automated base model is procedurally coherent, but it is also possible that it lacks the usual writing modality of the person who signs the order. To resolve this obstacle, the model can be asked to imitate the writing style of other offices and adapt the generated model to achieve similarity.

This correction can be considered the final step of the automation process for generating draft dispatches, so it is the right time to carry out all the corrections and supervisions inherent to the professional work.

Below the model *prompt* style appropriateness:

Based on this first text [decree/dispatch/providence model with the style of the person signing the dispatches] rewrite the second text [automated base decree/dispatch/providence model] imitating the way of writing of the first text."

v. For the approach, analysis and resolution of legal and non-legal problems

v.1.*Super Prompting* for exploring diverse approaches to a problem[16]

It is a technique of *prompting* that matches multidisciplinary logic with a response approach based on *Tree of Thoughts*.

It is done through a *super prompt* that contains several commands or indications together that optimize the dialogue with the LLM. To do this, it proposes the intervention and response, in a cascade and jointly, of more than one expert in one or several topics (there may be two or more iterations).

This interaction methodology allows improving the exploration of coherent text units (thoughts) that serve as intermediate steps for problem solving; optimizes the human/AI dialogue, since it allows generating a more complete response (the AI assumes three roles at the same time and confronts them in deliberation); expands the possibilities of thinking about a phenomenon based on the "simulated dialogue" between the predictions and the generation of synthetic data behind each "expert"; exponentially increases productivity in the face of complex tasks such as the development and analysis of process maps, task maps, decision trees and conceptual matrices based on various perspectives; optimizes the probabilistic approach, the representation and simulation of a part of human knowledge provided by generative AI.

---

[16] Expand in Juan Gustavo Corvalan, "A Superprompt to enhance human intelligence. How to get 3, 6 or more "experts" to improve generative AI responses?", post from June 24, 2023, inhttps://www.linkedin.com/posts/juan-gustavo-corvalan_un-superprompt-para-potenciar-la-inteligencia-activity-7078132462513528832-1cXP/?utm_source=share&utm_medium=member_ios [accessed 6/24/2023].

Tips for using the *super prompt*:

- Clarify and segment orders in the same *super prompt*
- Respect punctuation rules
- In the first *prompt* include all the orders and anticipate that the topic will be given in a second *prompt*
- In a second *prompt* introduce the theme
- Use the "conceptual matrix" formula
- During the interaction, if the model releases the fulfillment of the proposed task, indicate "ok, continue."

Below is a basic model of *super prompt* prepared from UBA IALAB:

Hello ChatGPT. I need you to act as three experts in ____, ____ and ____. They must debate the problem or issue that I will indicate to you. The debate must be iterated and I ask you to reflect the arguments in a table. I need you to then reach an agreed conclusion with a proposal or recommendation and to develop a conceptual matrix. Then I'm going to ask three other experts in ____, ____ and ____ to put the arguments and conclusion into crisis. Below I write the problem or question...

v.2. Strategies for evaluating evidence and seeking suggestions for resolving cases

Text generative AI can also be used as an assistant for evaluating evidence and facts to obtain suggestions or alternatives for resolving cases; for help to detect inconsistencies or contradictions in the evidence and facts presented; to evaluate argumentative coherence and identify inconsistencies in the arguments presented; to identify key points of the facts of a case and draw conclusions and also to identify and generate counter arguments or alternative points of view in order to consider different perspectives on the topic under analysis.

Specific suggestions for evaluating evidence and possible determination of the resolution of a case

1. First of all, there is a *prompt* through which the model is informed that we request its help to make a decision in a complex case. We outline a brief summary of the case and indicate that in a subsequent prompt we will inform

of the evidence and facts. The purpose of the provided texts is to obtain the help of ChatGPT for the evaluation of evidence and facts.

a. A first model is provided below *prompt*:

Hi GPT-4. I hope you are fine. I need your help. I am a judge [insert legal specialty] and I have to make a decision in a complex case that I have to deal with in the court that I am in charge of. The case [insert text summarily describing the facts of the case]. I will discuss the evidence and facts in more detail in the next prompt. Could you help me evaluate the evidence and facts so that I can determine [insert text explaining the intended objective. For example, if the dismissal was due to a cause attributable to the plaintiff or was it due to a cause attributable to the defendant who fired her due to the pregnancy]?

b. Then we proceed to provide the model with a second *prompt* which details the facts and evidence.

[insert text that describes the events in sufficient detail]. As a judge I must evaluate these facts and evidence to decide [indicate what decision you should make to provide context. For example, if the dismissal was for a cause attributable to the plaintiff or it was for a cause attributable to the defendant who dismissed her due to the pregnancy]. Can you help me?

2. Prompt model to carry out an analysis of coherence and consistency between evidentiary means. For example, between testimonies.

Hello ChatGPT. I ask you to review and evaluate the logical and argumentative coherence between the testimonies A, B and C that I am going to provide you. Identify points of agreement and points of contradiction between them. I ask you to put them in a box. Below I give you the testimonies: [insert the text with the testimonies of A, B and C].

vi. To work on legal and non-legal texts

vi.1. Text simplification

Text generative artificial intelligence can be very useful to simplify technical language and generate a text expressed in more accessible terms (clear language) and/or free of masculinities (inclusive language), with the aim of improving and facilitating communication with the citizens.

Models of *prompt* to simplify a text and propose an output in clear and/or inclusive language:

1. *I need to simplify the following paragraph. Use legal language, keep data that refers to law numbers, deadlines, amounts of money, names of entities* data *or people*: *[insert the text you want to simplify].*

2. *I need you to explain to me the text that I am going to provide you below, and that you do so using clear and accessible language for a non-specialized audience. I ask you to maintain the formal tone, preserve the legal language as much as possible and bring warmth to the writing: [insert the text you want to simplify].*

3. *I need you to explain to me the text that I am going to provide you below, and that you do so using clear and accessible language for a non-specialized audience: [insert the text you want to simplify].*

4. *I need you to write this text eliminating masculinities, while also using clear and accessible language for a non-specialized audience. I ask you to maintain the formal tone, preserve the legal language as much as possible and bring warmth to the writing: [insert the text you want to simplify].*

vi.2. Text summary

Text generative artificial intelligence is useful for summarizing legal and non-legal texts, because it allows this task to be carried out with good precision and very quickly. In this way, the model can be required to identify and extract relevant information, find keywords, and discard unimportant content.

Two models of *prompts* are provided below to summarize a non-legal and legal text:

1. *Hello ChatGPT. I need you to summarize the key points of the text that I am going to provide below. You must also provide me with the conclusions: [insert the text you want to summarize].*

2. *Hello ChatGPT. I need you to summarize the key points of the text that I am going to provide below. I also ask that you identify the relevant facts and legal arguments, and provide me with conclusions: [insert the legal text you want to summarize].*

VII. Team:

General Management: Juan Gustavo Corvalán and Mariana Sánchez Caparrós

UBA-IALAB research team: Giselle Helleg, Carina Papini, Melisa Raban and Antonella Stringhini.

Dyntec-Lab research team: Marco Rossi, Franco Orellana.

VIII. Do you want to be part of the Guide? Scan the QR and leave us your contributions

From UBA-IALAB we want you to be part of the construction of this Guide.

To do this, we enable this participatory, open and transparent space with which we begin the public consultation so that you can make your suggestions and propose new uses, models of *prompts*, guidelines and recommendations not included in this first version, that you understand are relevant to meeting the objectives of the document.

You can send your contributions and opinions through this form.

Thank you very much for being part of this project!