**31.** [**Guidelines for the Use of Artificial Intelligence in University Courses**](https://forogpp.files.wordpress.com/2023/02/guidelines-for-the-use-of-artificial-intelligence-in-university-courses-v4.3.pdf)

Guidelines for the Use of Artificial Intelligence in University Courses Prepared by Juan David Gutiérrez, Universidad del Rosario1 Version 4.3 (21/02/2023). License C.C. BY 4.0

**1. Justification:**

* We use artificial intelligence tools daily. Consciously or unconsciously, practically all of us use some type of artificial intelligence (AI) in our daily lives. We even use systems that help us perform various university activities. For example: when we are writing a text message or an email and the ‘autocomplete’ tool suggests how to finish writing the sentence; when we search for information through Internet search engines that recommend keywords for the search; when we use word processing software that suggests grammatical corrections; and, when we use automated translation tools, among others. In addition, there are web and mobile AI tools that are freely accessible to anyone with an Internet connection and that allow us to generate text and audio-visual content.
* Some AI can support learning processes and professional activities. There are appropriate uses of such technologies which can be useful for learning and teaching at universities. Moreover, there are different uses for these tools at the professional level; for example, AI tools contribute to activities in all the stages of public policy processes.
* But certain uses of AI are risky, so their use must be informed, transparent, ethical, and responsible. AI tools are not suitable for every type of activity and certain types of uses can be counterproductive to the pedagogical process. Some uses of AI may generate risks for users and third parties. Furthermore, as explained in section 4 of the guidelines, it is essential to be aware of the ethical, environmental, and human rights implications associated with the use of these tools. It is precisely because of these risks that some universities have published recommendations and guidelines on the use of AI in academic contexts.

**2. Objectives:**

1. Prevention. To prevent situations in which students consciously or unconsciously engage in academic dishonesty.
2. Contribute to digital literacy. Contribute to learning basic knowledge about the use of emerging technologies such as AI, their benefits and risks, and their implications for society.
3. Promotion of responsible use. Promote the responsible and ethical use of emerging technologies, such as AI, in learning processes and for future use in professional life.

**3. Rules for the use of AI in and out of the classroom:**

1. General rule. In this course, the use of AI as a support tool to carry out different learning activities is allowed. The parameters for the use of these tools described below distinguish between

‘low risk’ and ‘high risk’ tools, according to the risks that the respective AI generates for the pedagogical process, for the users, and for third parties.

1. Use of low-risk AI. You may freely use low-risk AI tools, i.e., those that allow you to correct or review student-generated content or those that allow you to collect and process data. For example, grammar correction tools, translation tools, audio-to-text transcription tools, and Internet information search tools, among others. It is recommended that where such tools enable you to do meaningful work (e.g., translation), you indicate their use in the appropriate section (e.g., methodology section).
2. Use of high-risk AI. AI tools that generate content (text and/or audio-visual), i.e., textgenerating

AI (generative AI, such as ChatGPT) and stable diffusion AI for image generation (Stable Diffusion, such as DALL-E 2 or Midjourney) are considered high-risk in these guidelines. You may only include AI-generated content in your individual and group work when four requirements are fulfilled: a. Informed use. Prior to using the tool, research who or what company developed the tool, how it was developed, how it works, what functions it can perform, and what limitations and/or risks it presents. b. Transparent use. In your work, indicate in detail and expressly which tool you used and how you used it (a requirement currently requested by scientific journals to those who submit manuscripts for peer review). c. Ethical use. The manuscripts must distinguish what was written or produced directly by you and what was generated by an AI tool. On this point, general citation rules apply, e.g., use quotation marks if you include textual paragraphs. Violations of this policy, particularly with respect to the second requirement, will be considered an infringement to academic integrity. In addition, ChatGPT is not a person, so it cannot be considered your co-author (just as you should not include Google as your co-author). Some scientific journals have already updated their publication policies to clarify that language models such as ChatGPT do not satisfy their ‘authorship’ attribution criteria. d. Responsible use. It is recommended that the use of these AI tools be limited to early stages of research, to inspire or suggest directions, not to produce content that will later be included in your deliverables. In any case, if you choose to transcribe texts produced by Generative AI, you must prove that you have rigorously checked such information against reliable sources, since Large Language Models (LLMs) such as ChatGPT tend to offer inaccurate, erroneous, and invented information.

1. Monitoring. The teacher may use tools to detect violations associated with the use of highrisk tools. Turinitin announced that it is working on adding an AI handwriting detection module. In addition, it is likely that, soon, the generative tools themselves will include a ‘watermark’ in their results that will make it easier to track whether a text was produced by AI. In any case, teacher will focus on pedagogical strategies rather than punitive ones.
2. Accompaniment of the teacher. At the beginning of the semester, the professor will explain what AI is, what language models are and how tools such as ChatGPT work, what opportunities and risks they generate for academic and professional work, and what ethical, environmental, and human rights implications are associated with the use of these tools. The professor will always be available to clarify the scope of these guidelines, to discuss and co-create them, and to resolve pointed questions about the use of AI.

**4. Why is an informed, transparent, ethical, and responsible use of AI necessary?**

• The use of AI tools must be informed, transparent, ethical, and responsible for –at least– four types of reasons:

1. because AI tools are not always reliable;
2. because there are risks that certain uses will negatively affect learning processes;
3. because of the risk of users treating AI as if it were a human being; and,
4. because the use of the tools has ethical and human rights implications due to the way they were developed and/or because some tools may replicate or amplify social issues such as discrimination. Each of the four reasons is explained in detail below: o First, because their answers are NOT always reliable even though, for example, an AI generator produces text that looks convincing. LLM-based systems such as ChatGPT do not perform with the accuracy of other tools used in learning environments such as calculators. In fact, ChatGPT tends to include false or fanciful information in its responses. Microsoft's built-in chatbot for its Internet browser, Bing, and demonstrations of Google’s chatbot, Bard, have also presented the same types of problems. These systems do not distinguish true from false. Why does this happen? LLMs spin words from probabilistic inferences from the data they were trained on, but they do not have the ability to understand what they produce or associate meanings to the words they utter (they are ‘stochastic parrots‘). Recently, a media outlet that used a ChatGPT-type tool to write texts had to publish corrections to multiple articles due to serious inaccuracies. o Second, because the use of textgenerating tools may discourage students' motivation to write and think on their own. It is worth reiterating that the learning activities in this subject seek to develop their cognitive skills and that this policy seeks to prevent some AI tools from becoming automated plagiarism mechanisms. o Thirdly, because of the risk that users, consciously or unconsciously, treat AI behavior as if it were human (Eliza Effect). LLM-based tools do not understand their output, but simply mimic language patterns from the synthesis of large volumes of data from which the program generates word sequences. This problem of anthropomorphizing machinesis exacerbated by the fact that some chatbots have produced violent or harassing responses that could lead people to engage in harmful behavior. o Fourth, because of the ethical and human rights implications associated with the use of certain AI systems given that: some tools tend to reproduce or amplify derogatory and discriminatory stereotypes associated with gender, race, ethnicity or disability; technologies could have been developed from massive copyright infringement; new forms of colonialism through the non-consensual extraction of information from historically marginalized communities; some tools would have been developed in contexts of labor exploitation; the development and operation of such systems generates a considerable carbon footprint; and, the potential violation of privacy and personal data protection rights of those who use them.

**5. Open-Guidelines:**

• AI is a rapidly changing set of tools, which is why these guidelines will remain open to future evaluation, modification, and revision. From the beginning of the semester and throughout the semester the professor will open spaces to discuss the guidelines with the students and, if necessary, modifications to this guideline can be introduced through co-creation exercises. • For those interested in the challenges that generative AI creates for those of us who teach in universities, I recommend this document by Professors Anna Mills and Lauren M. E. Goodlad. • For those who would like to read a brief introduction on how LLMs work and what their main risks are, I recommend the article ‘On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? ’ by Emily M. Bender, Timnit Gebru, Angelina McMillan-Major and Shmargaret

Shmitchell. • Anyone interested in accessing material to learn basic information about AI produced by MIT, I recommend checking out the DAILy Curriculum project.

**32.** [**Guidelines for the use of ChatGPT and text generative AI in Justice**](https://ialab.com.ar/wp-content/uploads/2024/02/Guia-uso-IAG-.pdf)

#### 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) 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 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.

#### 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. 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 systems7 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.8In 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 finetuned 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 year11 , 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. 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 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.

1. 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

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. 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

1. 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 thirdparty 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 AIgenerated texts. - Establish multidisciplinary consultation teams so that they can address the study and work of complex use cases.

1. 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.

1. 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 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. 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

1. 1.Super Prompting for exploring diverse approaches to a problem

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. 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...

* 1. 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].

1. 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].

* 1. 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].

1. **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.

1. **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.

**33.** [**Guidelines for the use of artificial intelligence in university contexts**](https://juangutierrezco.files.wordpress.com/2023/08/guidelines-for-the-use-of-artificial-intelligence-in-university-contexts-v5.0.pdf)

#### Introduction

* We use artificial intelligence tools daily: Consciously or unconsciously, practically all of us use some kind of artificial intelligence (AI) in our daily lives, in particular systems based on language models. For example, we use language models when we are writing a text message or an email and the “autocomplete” tool suggests how to finish writing the sentence. We also use language models to support us in carrying out academic activities, such as automated translation systems or programs that detect grammatical errors, among others. In addition, nowadays, there are web platforms and mobile applications that provide access to AI tools to generate texts and audiovisual content
* Some uses of AI tools can support learning, research, teaching, and other professional activities: There are appropriate uses of such technologies, which can not only be useful to support learning, teaching, and research processes at the university but also at a professional level. For example, different AI tools nowadays contribute to activities at all stages of public policy processes, and, in Colombia, more than 50 public sector entities have adopted automated decision systems to support the performance of their functions.
* But certain uses of AI are risky, so their use must be informed, transparent, ethical, and responsible:AI tools are not suitable for every type of activity and certain types of use can be counterproductive to the pedagogical process. Some uses of AI can generate risks for users and third parties. Furthermore, as explained in section 4 of these guidelines, it is essential to be aware of the ethical, environmental, and human rights implications associated with the use of these tools. It is precisely because of these risks that some universities and professors have published guidelines, recommendations, principles, policies, and directives on the use of AI in academic contexts (see list in section 5 of these guidelines).
* We should not assume that all our students are “digital natives” or that they want to adopt new technologies: . People's attitudes towards technologies are diverse and complex, it should not be assumed that all our students are eager to adopt new technologies. For example, in the surveys I have carried out in workshops with students on the use of AI at university, I have found that these systems generate very diverse feelings in them: fear, intrigue, excitement, uncertainty, curiosity, distrust, euphoria, worry, amazement, expectation, anguish, tranquility, confusion, fascination, caution, questioning, rejection, uneasiness and comfort. Moreover, people's capacity to adopt new technologies is heterogeneous and depends on different economic, social and cultural factors, among others. All the above implies that those of us who teach should implement differentiated teaching strategies according to the needs of our students and that we should not force the use of new AI systems.
* Guidance for the use of AI inside and outside the university classroom: These guidelines provide basic guidance for university professors and students on how to use AI, particularly large-scale language models, in an informed, transparent, ethical, and responsible manner. The challenges of using AI in education are not limited to aspects associated with learning processes and academic research. For example, the adoption of AI to (semi)automate personnel hiring and evaluation processes or for the surveillance of educational facilities also has legal and ethical implications. But these issues are beyond the scope of these guidelines since the rules and principles outlined here focus exclusively on the use of AI for educational and academic purposes inside and outside the university classroom.
* Definitions of three key terms: The following are working definitions of three technological terms used in these guidelines: (A) Artificial Intelligence (AI) System: “a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions that influence real or virtual environments. AI systems are designed to operate with varying levels of autonomy”. (B) Language model: algorithms that process, analyze and produce natural language in written or spoken form. (C) Large Language Model (LLM): algorithms trained on a vast amount of information whose main function is to predict the most likely string of words given the context that precedes or follows it.

#### Objectives of these guidelines

•Contribute to digital literacy. Contribute to learning basic knowledge about the use of AI in university contexts – with emphasis on LLMs. This objective includes teaching about its benefits and risks, its implications for society, and the development of critical digital competencies.

* Promotion of responsible use of AI. Promote the responsible and ethical use of these technologies in learning processes and for future use in professional life.
* Prevent the violation of academic integrity. Prevent situations in which students consciously or unconsciously incur in academic dishonesty.

#### Rules for the use of AI in and out of the classroom

* General rule. The use of AI as a support tool to carry out different learning activities is allowed. The parameters for the use of these tools described below distinguish between “low risk” and “high risk” tools, depending on the risks that the respective AI generates for the pedagogical process, for users and for third parties.
* Rules for the use of low-risk AI. Students may freely use low-risk AI tools, i.e., those that allow to correct or review student-generated content or those that allow to collect and process data. For example, grammar correction tools, translation tools, audio-to-text transcription tools, and Internet information search tools, among others. Where such tools enable students to do meaningful work (e.g., translation), they must indicate their use in the appropriate section (e.g., methodology section). If students have doubts about how to classify the risk of using a specific AI, I suggest contacting their teacher.
* Rules for the use of high-risk AI. AI tools that are used to generate synthetic text are high-risk. That is, LLMs that can be accessed through chatbots, such as ChatGPT, Bard, Bing, and Claude, etc. Students can include AI-generated content in their individual and collective deliverables when four requirements are met: i. Informed use. Prior to using the tool, students must investigate who or what company developed the tool, how it was developed, how it works, what functions it can perform, what its terms of use are, what the company does with the information it collects from the user, and what limitations and/or risks it may generate for the user and/or third parties. ii. Transparent use. Students must explain in detail and expressly which tool were used and how the tools were used in the methodology section or in the section that they deem pertinent. This is a requirement currently requested by scientific journals and publishing houses, such as Nature and Cambridge University Press, to whom manuscripts are submitted for peer review. iii. Ethical use. In relation to the use of LLMs, students must distinguish what was written or produced directly by them and what was generated by an AI tool. On this point, general citation rules apply, e.g., use quotation marks if textual paragraphs are included. For citing synthetic text produced by LLMs, students may consult the rules suggested by the APA style team. Violations of this rule, particularly with respect to the second requirement, will be dealt with under an approach like that which applies in cases of plagiarism. In addition, ChatGPT is not a person, so it cannot be considered a co-author. Some scientific journals have already updated their publication policies to clarify that language models such as ChatGPT do not satisfy their “authorship” attribution criteria. On the other hand, the ethical use of AI systems – in general, not just LLMs – should be informed by four principles of bioethics: beneficence (achieving benefits), respect for the autonomy of persons (including obtaining consent if third parties are involved), justice (especially nondiscrimination), and nonmaleficence (not intentionally or recklessly causing harm to others). iv. Responsible use. The use of AI tools should be limited to early stages of research, to inspire or suggest directions, not to produce content that will later be included in students’ deliverables. In any case, if students choose to transcribe texts produced by an LLM they must prove that the content was rigorously checked with reliable sources, since these systems tend to offer inaccurate, erroneous and false information. Responsible use also implies avoiding that the use of the system does not cause damage to the user or third parties. For example, it is not recommended to include personal data or confidential information about the students or others in chatbot queries because once such information is entered, the user loses control of who can access it and there is a risk that it will be known by unauthorized third parties.
* Plausible uses of LLMs: i. Explore. To explore new issues and seek inspiration, i.e., use the chatbot in the earliest stages of research. ii. Automate basic tasks. To support exercises that involve recognition and/or reproduction of patterns in texts, for example: translating, summarizing, modifying the tone and style of the text – for example, to make the text clearer or more accessible to certain audiences –, correcting spelling and grammar, audio-text conversion, programming, etc. Students should no rely on automated tasks, the LLM's output must be carefully reviewed. iii. Trivial writing. To support unimportant writing tasks. To decide whether something is trivial or not, think about the following heuristic: “what would happen if my interlocutor found out that I used an LLM to write the message?”. iv. Sparring. Testing one's own arguments or ideas through interaction with the chatbot, for example, by asking the system to respond with criticisms or counterarguments to the student's statements.
* Not recommended uses of LLMs: i. Use the chatbot as the main or only source to search for factual or technical information. I do not recommend it, LLMs are not reliable search tools. Besides, Internet search engines are more efficient and reliable for that task. In any case, since searching chatbots can be perceived as an easier and faster way to obtain information, I recommend that the information obtained through chatbots should always be carefully checked with reliable sources. ii. Use of synthetic text for substantive issues. I do not recommend it because of the tendency of LLMs to produce answers with inaccurate, erroneous and false information. In any case, if students do so, each point should be checked against other reliable sources. iii. Doing mathematical calculations. LLMs are not trained to perform mathematical calculations and thus often fail to answer simple arithmetic operations.
* Accompaniment of the teacher and development of digital competencies. The teacher will dedicate class time to explain what AI is, what are LLMs that power tools like ChatGPT, what opportunities and risks do these systems generate for academic and professional work, and what are the ethical, environmental, and human rights implications associated with the use of these tools. The professor will be available to clarify the scope of these guidelines, to discuss and co-create them, and to resolve specific questions about the use of AI. In addition, the teacher will seek to carry out activities that contribute to the digital literacy of students, particularly for the development and strengthening of digital competencies such as the ability to critically evaluate the benefits and risks of new technologies.
* Supervision. The teacher will use different sources to identify potential situations of academic dishonesty using AI, but he or she will not rely on the tools available to detect synthetic text since none of these tools are currently reliable. In any case to avoid risks associated with academic integrity, the main strategy the teacher will use will be pedagogical not punitive.

**Why the informed, transparent, ethical, and responsible use of AI in academia is necessary?**

* The use of AI tools must be informed, transparent, ethical and responsible for – at least – four types of reasons: (1) because LLMs are not always reliable; (2) because there are risks that certain uses may negatively affect learning processes; (3) because of the risk of users treating AI as if it were a human being; and (4) because the use of the tools has ethical and human rights implications due to the way they were developed and/or because some tools may replicate or amplify social issues such as discrimination. Each of the four reasons is explained in detail below: • First, because their answers are NOT always reliable even though, for example, an AI generator produces text that looks convincing. LLM-based systems such as ChatGPT do not perform with the accuracy of other tools used in learning environments such as calculators. In fact, ChatGPT tends to include false or fanciful information in its responses. Microsoft's built-in chatbot for its Internet browser, Bing, and Google's chatbot, Bard, have also presented the same types of problems. These systems do not distinguish true from false. Why does this happen? LLMs spin words from probabilistic inferences from the data they were trained on, but they do not have the ability to understand what they produce, nor do they associate meanings to the words they output (they are “stochastic parrots”). • Second, because the use of text-generating tools may discourage students' motivation to write and think on their own. It is worth reiterating that the learning activities in this subject seek to develop their cognitive skills and that these guidelines seek to prevent some AI tools from becoming automated plagiarism mechanisms. • Thirdly, because of the risk that users, consciously or unconsciously, treat AI behavior as if it were human (Eliza Effect). For example, LLM-based tools do not "understand" the texts they produce; they mimic language patterns by synthesizing large volumes of data based on which they generate word sequences. This problem of anthropomorphizing machines is exacerbated by the fact that some chatbots have produced violent or harassing responses that could lead people to engage in harmful behavior. • Fourth, because of the ethical and human rights implications associated with the use of certain AI systems given that: some tools tend to reproduce or amplify derogatory and discriminatory stereotypes associated with gender, race, ethnicity or disability; technologies could have been developed from massive copyright infringement; new forms of colonialism through the non-consensual extraction of information from historically marginalized communities; some tools would have been developed in contexts of labor exploitation; the development and operation of such systems generates a considerable carbon footprint; and, the potential violation of privacy and personal data protection rights of those who use them. For more information on the harms associated with the use of Generative AI, I recommend consulting the report “Ghost in the machine: Addressing the consumer harms of generative AI” (2023) by the Norwegian Consumer Council and the report “Generating Harms: Generative AI's Impact & Paths Forward” by the Electronic Privacy Information Center

#### Other resources for teachers

• Other guidelines on the use of AI at the university. I recommend consulting guidelines, policies, and principles of other universities, professors, and other organizations, such as: University of Washington, Hertie School, University of Tartu, Yale University, University of Helsinki, Technical University of Munich, University College London, UNESCO, UK Department of Education, Boston University, Warwick University, Universidad de Buenos Aires (IALAB),

Universidad del Desarrollo, Russell Group, and Lance Eaton's list. • Introductory literature on

LLMs. (A) For an introduction on how LLMs were developed and how they work (in particular, GPT-3), I recommend this text by Timothy B. Lee and Sean Trott. (B) For an explanation on how

LLMs work and what are their main risks, I recommend the text “On the Dangers of Stochastic

Parrots: Can Language Models Be Too Big?” by Emily M. Bender, Timnit Gebru, Angelina

McMillan-Major and Shmargaret Shmitchell. • Teaching resources. (A) For those interested in the challenges that LLMs create for teaching writing in universities, I suggest consulting the paper entitled “Adapting College Writing for the Age of Large Language Models such as ChatGPT: Some Next Steps for Educators” by Anna Mills and Lauren M. E. Goodlad. (B) I also suggest the materials contained in “AI Text Generators and Teaching Writing: Starting Points for Inquiry” by Anna Mills. (C) The article by Anna Mills, Maha Bali and Lance Eaton entitled “How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process” documents and reflects on how professors from different universities around the world have adopted new pedagogical practices to respond to the challenges generated by Generative AI. (D) MIT offers for free its DAILy Curriculum program that contains materials to teach basic AI skills to teenagers. • Critical approach to the use of AI in academic settings. (A) I recommend consulting Maha Bali's list of resources on developing critical literacy in AI. (B) I also suggest reading a reflection by Mohammad Hosseini, Lex Bouter, and Kristi Holmes in which they argue that it is not desirable to adopt AI tools in education without first reflecting on their effects and biases and without having adopted measures to mitigate their risks. (C) Naomi S. Baron's article “How ChatGPT robs students of motivation to write and think for themselves” discusses how the use of LLMs can negatively affect students' creative processes. (D) Dimitrinka Atanasova's blog post reflects on how LLMs can reduce barriers for students and researchers to write in English (when this is not their first language), but that their effects on cultural inclusivity in higher education are more ambiguous. (E) Finally, I recommend the Kathryn Conrad's text “A

Blueprint for an AI Bill of Rights for Education” that proposes rights for teachers and students that can inspire policy formulation and protective measures.

#### Guidelines open to change

• AI is a rapidly changing set of tools, which is why these guidelines will remain open to future evaluations, modifications, and revisions. • Throughout the academic semester, the professor will open spaces to discuss these guidelines with the students and, if necessary, modifications may be introduced through co-creation exercises. • I am grateful to my colleagues and others who provided comments and criticisms on initial versions of this document and also to those who have participated in the talks and workshops I have given on the use of large-scale language models in university contexts.

#### Changes introduced in this version of the guideline

• I introduced changes of form and substance with respect to the previous version of this instrument (v. 4.3), which I summarize below. • Changes of form: (A) The previous version was called “Guidelines for the Use of Artificial Intelligence in University Courses”, the new title is broader since the instrument is useful for academic areas that go beyond classroom activities (e.g., research activities). (B) I introduced new sources that can be consulted in the form of hyperlinks throughout the text. • Substantive changes: (A) In point 1 of the new version of the guidelines I included a glossary of key terms and in point 3 I introduced a new sub-section with plausible and nonrecommended uses of LLMs by students. (B) In addition, in this version I included a new section of resources for teachers that includes examples from other guidelines, basic literature on LLMs, resources for teaching, and critical literature on the use of AI in educational contexts. (C) On the other hand, the previous version of this instrument indicated that the teacher would use synthetic text detection systems, but I have removed those allusions in the present version because the currently available systems produce a high level of false positives and negatives. For this reason, for example, the Teaching Center at the University of Pittsburgh issued a press release stating that it did not endorse or promote any of the tools currently available

#### Use of these guidelines in your classes

• These guidelines are published under a Creative Commons Attribution 4.0 license, so if you are an educator interested in formulating your own policies or guidelines, you can share and adapt these guidelines as long as you make the corresponding attribution and indicate if you have made any changes. • If you would like to send me your comments, criticisms, and suggestions, please write to me at juagutie@uniandes.edu.co

**34.** [**Editorial Policy, Publication Ethics and Malpractice Statement**](https://revistas.javeriana.edu.co/files-articulos/CRC-EPUJ/manuales/ETHICS/)

#### Editorial Policy

The scientific and academic journal system of Pontificia Universidad Javeriana is composed of a set of indexed periodic publications managed jointly by the editorial teams of the Faculties and Editorial Javeriana. We are committed to open and transparent communication, scientific integrity and ethical standards in all aspects of the editorial and dissemination process. Our editorial policy is governed by the following principles:

*Quality:* It is our main goal to publish high-quality research that is relevant, visible and impactful. All submissions are subjected to rigorous peer-review by independent experts in the field. We strive to provide constructive feedback to authors to help them improve their work.

*Ethics:* We are committed to upholding the highest standards of ethical conduct in our editorial process. We require authors to adhere to ethical guidelines for research, informed consent, protecting privacy and confidentiality, and avoiding plagiarism and fabrication of data. Every paper is checked by means of anti-plagiarism and AI-generated text software detection.

*Openness:* We believe in open and transparent communication and support open access to scientific knowledge. We encourage authors to share their data and code whenever possible, and we require authors to disclose any potential conflicts of interest.

*Diversity and inclusion:* We recognize the importance of diversity and inclusion in research and editorial processes. We strive to ensure that our journals represent a broad range of origins, perspectives and experiences, and we are committed to fostering a culture of respect, inclusivity, and equity in our editorial policies and practices.

*Innovation:* Innovative and unconventional approaches to research and policy analysis are welcome. We encourage authors to challenge existing paradigms and propose new ideas for Latin America and the world. We offer traceability of the impacts of research in academia, industry and society. It is possible to include podcast, video, animations, or interactive content in the HTML versions of the articles.

*Engagement:* Seeking to engage with our readers, authors, reviewers, and stakeholders in an ongoing dialogue about the most pressing issues in Latin America and the world, we encourage authors to communicate their research findings in a clear and accessible way, providing a forum for discussion and debate around important issues.

*Sustainability:* As open access electronic journals, sustainability is promoted in all aspects of our operations. We recognize that our activities have an impact on the environment, and we strive to minimize this impact by adopting sustainable practices:

* Energy Consumption: minimizing our energy consumption by implementing energyefficient practices in our office space, server rooms, and data centers is one of our main goals. We encourage our staff to switch off computers and other electronic devices when not in use. We use energy-efficient equipment wherever possible.
* Paperless Communication: We strive to operate as a paperless office by minimizing the use of paper and encouraging the use of digital communication, storage, and archiving methods. We also minimize printing and use recycled paper when printing is necessary.
* Green Hosting: We use web hosting services that are powered by renewable energy and strive to minimize the environmental impact of our online presence. We also optimize our website and use selected directories, databases and scholarly indexes to reduce our carbon footprint.

#### Copyright and Licensing

The journals published by Pontificia Universidad Javeriana are registered under Creative Commons Attribution 4.0 International Public License. Thus, all articles may be reproduced, distributed, and publicly shared in digital format, as long as the names of the authors and Pontificia Universidad Javeriana are acknowledged. Others are allowed to quote, adapt, transform, autoarchive, republish, and create based on this material, for any purpose (even commercial ones), provided the authorship is duly acknowledged, a link to the original work is provided, and it is specified if changes have been made. Pontificia Universidad Javeriana does not hold the rights of published works and the authors are solely responsible for the contents of their works; they keep the moral, intellectual, privacy, and publicity rights.

Approving the intervention of the work (review, copy-editing, translation, layout) and the following outreach, is granted through an use license and not through copyright transfer.

Publishing in this journal system does not generate direct royalties for contributors.

#### Privacy Statement

The names and email addresses submitted to the journals edited by Pontificia Universidad Javeriana will be used for the purposes stated solely and will not be made available for any other purpose or to any other party.

In accordance with Colombian Statutory Law 1581 of 2012 on Data Protection and related regulations, the Owner is informed that, by voluntarily providing and registering their data through the channels enabled on this website, they authorize their incorporation into a database under the responsibility of the Pontificia Universidad Javeriana, to be treated for the purpose of carrying out associative, cultural, recreational, sports, and social activities, management of social media and/or editorial content, historical, scientific or statistical purposes, administrative procedures and publications, and accounting, tax, administrative and advertising management and commercial prospecting - Own advertising. Likewise, I authorize my personal images to be published in printed and audiovisual media, institutional websites and social networks.

It is optional to provide information about Sensitive Data, understood as those that affect privacy or generate any type of discrimination, or about minors.

The policy on the treatment of the Owner's data, as well as substantial changes that occur in it, can be consulted through the following email address: usodedatos@javeriana.edu.co. Likewise, it will be kept updated on the entity's website: [Agreement 657 (in Spanish)](http://www.javeriana.edu.co/documents/10179/9567882/Acuerdo+No+657.pdf)

You may exercise your rights of access, correction, deletion, revocation, or claim for infringement of data, by sending a written request to Pontificia Universidad Javeriana, to the email address usodedatos@javeriana.edu.co, indicating in the subject the right you wish to exercise, or by ordinary mail sent to the address: Carrera 7 # 40-62 in the City of Bogota, Colombia.

#### No Author Fees

Our Diamond Apen Access journal system is completely fee-free for authors. We believe that access to knowledge should not be limited by financial barriers, and we are committed to making our journals accessible to researchers and scholars from all backgrounds and regions.

We do not charge any submission fees, article processing fees, or publication fees. All articles published in our journals are freely available to readers worldwide, ensuring that research is disseminated widely and can be accessed by anyone with an internet connection.

Our commitment to no fees does not compromise the quality of our publication. We employ rigorous peer-review processes to ensure the integrity and validity of the research we publish. Our editors and reviewers are experts in their fields and uphold the highest standards of academic excellence.

We welcome submissions from all researchers, regardless of their funding or institutional affiliations. By eliminating financial barriers, we hope to encourage a more diverse and inclusive academic community and foster the growth of new ideas and perspectives.

#### Publication Ethics Statement

In order to maintain the highest ethical standards in all aspects of our operations, we have established the following ethics statement to guide the behavior of our editorial team, authors, and reviewers:

***ETHICAL STANDARDS FOR EDITORS***

Our editors are responsible for ensuring that all published content adheres to the highest ethical standards. As such, they will ensure that:

* The editorial decision-making process is objective, transparent, and free from any commercial or personal interests.
* Any conflicts of interest, whether actual or perceived, are identified and disclosed to the relevant parties.
* All submitted manuscripts are treated confidentially and any information related to the manuscript is kept confidential.
* Any suspected ethical misconduct or malpractice is investigated thoroughly, and appropriate actions are taken.
* The editorial team takes measures to ensure that any published research involving human subjects, animals, or potentially hazardous materials adheres to relevant ethical guidelines.

***ETHICAL STANDARDS FOR AUTHORS***

Authors are responsible for ensuring that their research adheres to the highest ethical standards.

As such, they will ensure that:

* The research presented in the manuscript is original and has not been published elsewhere.
* The manuscript is free from plagiarism, falsification, or fabrication of data.
* Any potential conflicts of interest are disclosed in the manuscript.
* Any research involving human subjects, animals, or potentially hazardous materials adheres to relevant ethical guidelines, and the necessary permissions have been obtained.

The manuscript is written in clear and concise language, and the research is presented in an accurate and unbiased manner.

***ETHICAL STANDARDS FOR REVIEWERS***

Reviewers play a critical role in maintaining the quality and integrity of our journals. As such, they will ensure that:

* The review process is objective, transparent, and free from any commercial or personal interests.
* Any conflicts of interest, whether actual or perceived, are identified and disclosed to the editorial team.
* Any suspected ethical misconduct or malpractice is reported to the editorial team.
* The review is conducted in a timely and constructive manner, and any criticisms or suggestions are presented in a respectful and professional manner.

#### COPE adherence Statement

Our scholarly journals are committed to upholding the highest ethical standards in publishing and adhere to the guidelines and best practices set forth by the Committee on Publication Ethics (COPE):

* Editorial Independence: Our editorial decisions are independent and free from undue influence, whether from owners, sponsors, advertisers, or other third-party entities.
* Peer Review: Our peer review process is transparent and impartial, and all manuscripts are subject to rigorous review by qualified experts in the field.
* Authorship and Contributorship: We follow the Contributorship Defined (CreDIT) model for authorship, which acknowledges the contributions of all individuals involved in the research process.
* Conflict of Interest: We require authors, reviewers, and editors to disclose any potential conflicts of interest that may compromise the integrity of the research or publication.
* Research Misconduct: We do not tolerate research misconduct, including plagiarism, fabrication, and falsification of data.

Data and Reproducibility: We require authors to provide all relevant data and materials necessary to reproduce their findings and support their conclusions.

* Intellectual Property: We respect the intellectual property rights of authors and require authors to obtain appropriate permissions and/or licenses for any copyrighted material used in their manuscript.
* Post-publication Discussions and Corrections: We provide a platform for post-publication discussions and corrections, and we are committed to correcting any errors or inaccuracies in our publications.

By adhering to the guidelines and best practices set forth by COPE, we are committed to promoting transparency, integrity, and responsible publishing practices in all aspects of our journals. We encourage our authors, reviewers, and readers to report any concerns or potential breaches of these practices to our editorial team for review and action.

***6.1. AUTHORSHIP AND CONTRIBUTORSHIP***

Our scholarly journals follow the Contributorship Defined (CreDIT) model for authorship, which acknowledges the contributions of all individuals involved in the research process. The CreDIT model promotes transparency and accountability in authorship and helps ensure that all contributors are recognized for their contributions.

In accordance with the CreDIT model, authors must explicitly state their individual contributions to the research presented in the manuscript. The contributions may include, but are not limited to, the following:

* Conceptualization: This includes the formulation of research questions, study design, and hypothesis development.
* Methodology: This includes the development or adaptation of methodologies and protocols for data collection and analysis.
* Data Curation: This includes the management and organization of data, including data entry, quality control, and data archiving.
* Formal Analysis: This includes the application of statistical methods and techniques to analyze the data.

Investigation: This includes the collection and acquisition of data and/or samples.

* Resources: This includes the provision of funding, facilities, equipment, and other resources necessary for the research.
* Writing - Original Draft: This includes the initial drafting of the manuscript, including the writing of the introduction, methods, and results sections.
* Writing - Review & Editing: This includes the critical review and editing of the manuscript, including the revision of the introduction, methods, and results sections.
* Visualization: This includes the creation of visual aids, figures, and tables to present the data.
* Supervision: This includes the oversight and guidance of the research team and/or project.
* Project Administration: This includes the coordination and management of the research project, including obtaining ethical approvals, obtaining funding, and other administrative tasks.

Each author must provide a statement outlining their individual contributions to the research presented in the manuscript using the above categories. All authors must agree on the final author list and order of authors.

Our journal system recognizes that authorship credit should be based on substantial contributions to the research and does not allow honorary or gift authorship. We also encourage authors to acknowledge the contributions of individuals who do not meet the criteria for authorship but have contributed to the research in other ways.

By following the CreDIT model for authorship, we aim to promote transparency and accountability in authorship and recognize the contributions of all individuals involved in the research process.

***6.2. ARTIFICIAL INTELLIGENCE GENERATED TEXT AND IMAGES***

Artificial Intelligence (AI) generated text and images are not prohibited, but must be used in a responsible and ethical manner. All users of these technologies, like ChatGPT or Midjourney must adhere to the following principles:

Respect for Human Dignity and Rights: All text generated text and images must respect the dignity and rights of all individuals, regardless of their race, gender, sexual orientation, religion, or any other characteristic. Hate speech, discrimination, or any other form of disrespect towards individuals or groups is strictly prohibited.

* Accuracy and Fact-Checking: Users must take responsibility for verifying the accuracy of any text generated before relying on it. Users must take steps to ensure that the generated text is accurate and reliable.
* Avoiding Harm: Users must ensure that any text generated does not cause harm to individuals or society as a whole. This includes avoiding the dissemination of false or misleading information that could lead to harm, as well as avoiding the promotion of harmful behaviors or activities.
* Transparency: Users must be transparent about the use of Artificial Intelligence and any text or images generated by the model. This includes acknowledging the use of the software when publishing or sharing AI generated text or images, and providing clear explanations of how the content was generated.

AI software attributions must not appear as author or co-authors of any submission, nor the bibliographic reference list, but in the Methodology and/or Acknowledgements sections.

***6.3. ORIGINALITY AND PRIOR PUBLICATION***

The manuscripts submitted to this Journal System should be original. Authors must appropriately cite and attribute any work that has influenced their current contribution. This includes both published research articles and unpublished work such as conference presentations, theses, dissertations, multimedia or alternative sources.

Submissions should have not been previously published or submitted for publication elsewhere, in whole or in part in any language. Authors should declare that the manuscript is not a duplicate of any previously published work.

We do not consider posting on a preprint server and theses or graduation papers on institutional repositories (green-access portals) to be duplicate publication, because these typologies are not evaluated by peer-review process and have not received editoral or publishing intervention.

*Preprints*

* If there is a preprint version, authors must cite it in the new manuscript submitted to the journal (including Digital Object Identifier DOI and/or URL address) to ensure there will be a link to the original record.
* After publication, the following text can be added to preprint versions to encourage readers to use and cite the final published version of the article:

"This is an original manuscript of an article published in [JOURNAL TITLE], available online:

https://revistas.javeriana.edu.co/ | [Article DOI].".

All submitted manuscripts will be checked for originality using plagiarism detection software. Any text symilarity detected above 30 % may lead to further evaluation to determine if there is text recycling or any other malpractice.

***6.4. CONFLICTS OF INTEREST***

To uphold the highest standards of integrity and ethical conduct in all aspects of our publishing activities, all editors, reviewers, and authors are required to disclose any potential conflicts of interest that may arise during the publication process.

Conflicts of interest can arise in a number of different forms. For example, an editor may have a financial or personal relationship with an author or organization whose work they are evaluating. Similarly, a reviewer may have a personal or professional connection to an author that could influence their assessment of the manuscript.

To ensure that conflicts of interest are identified and appropriately managed, we require all editors, reviewers, and authors to disclose any relevant financial or non-financial relationships that could be perceived as having the potential to influence their judgment. This includes any relationships with organizations, companies, or individuals that may have a direct or indirect interest in the research or findings presented in the manuscript.

In cases where a potential conflict of interest is identified, we will take appropriate steps to manage the conflict and ensure that the integrity of the publication process is maintained. This may include assigning a different editor or reviewer to the manuscript, or requiring additional disclosures from the authors.

***6.5. DATA SHARING AND REPRODUCIBILITY***

*Data Sharing Policy:* authors are encouraged to make all data underlying their research available to other researchers, either through open data repositories or by providing the data upon request.

*Data Accessibility:* Authors should provide information on how to access their data, including any necessary software or programming code required to reproduce their results. The data should be accessible and usable by other researchers, and should be accompanied by clear documentation that describes the data and its provenance./p>

*Data Citation:* Authors should cite all datasets used in their research, and should provide complete references to these datasets in the manuscript. Data should be considered a valuable scholarly output and should be cited and recognized as such.

*Reproducibility:* Authors should provide all necessary information to enable other researchers to reproduce their research, including detailed descriptions of their methods, software code, and data sources. Authors should also include detailed descriptions of any data preprocessing, analysis, or interpretation performed in their research.

*Quality Control:* Authors should provide information on any quality control procedures used in their research, including how they identified and dealt with outliers, missing data, or other issues that may affect the validity of their results.

*Open Science:* Open science practices are recommended, such as pre-registration of research protocols and the use of open-source software tools. Open science practices can help to increase transparency, reproducibility, and accountability in research.

***6.6. MALPRACTICE STATEMENT***

To maintain the integrity of the research process and ensure high quality in every published work, scholarly journals must address various malpractices that can occur during the publication process. Some of the most common malpractices that scholarly journals encounter are plagiarism, salami slicing, double submission, and data fabrication/falsification:

* *Plagiarism* occurs when an author uses someone else's work, ideas, or words without proper attribution or citation. This includes copying and pasting text, paraphrasing without proper citation, or presenting someone else's work as one's own. Plagiarism undermines the credibility of the research and violates ethical and professional standards.
* *Salami Slicing* is the practice of breaking up one study into multiple smaller studies to increase the number of publications. This practice can lead to the duplication of data and ideas, and can also result in incomplete or misleading information being published.
* *Double Submission* occurs when an author submits the same manuscript to more than one journal at the same time. This practice can waste the time and resources of multiple journals, as well as delay the publication of other research.
* *Data Fabrication/Falsification* occurs when an author knowingly or unknowingly fabricates or falsifies research data or results. This can include manipulating data to obtain desired results, or fabricating data that was never collected or analyzed. Data fabrication/falsification undermines the credibility of the research and can have serious consequences for the scientific community and the public.

Other malpractices that can occur in scholarly publishing include conflicts of interest, duplicate publication, citation manipulation, and guest authorship. All of these malpractices can have serious consequences for the credibility of the research, the reputation of the authors, and the integrity of the scientific community.

***6.7. ALLEGATIONS OF RESEARCH MISCONDUCT***

In the event that ethical misconduct or malpractice is suspected, the editorial team will conduct a thorough investigation. If the investigation confirms the misconduct or malpractice, appropriate actions will be taken, including, but not limited to:

* Contacting the relevant authorities, such as the author's institution or funding agency.
* Retracting the published article.
* Banning the author from submitting to our journal system.

Removing the author from the editorial board.

We take any suspected ethical misconduct or malpractice very seriously, and we will take all necessary steps to ensure that the integrity of our journals is maintained.

***6.8. CORRECTION AND RETRACTION GUIDELINES***

Here are some general guidelines for authors who need to correct or retract a scholarly journal article:

* Act quickly and responsibly: If you discover errors or misconduct in your article, it is important to act quickly and responsibly. Delaying correction or retraction can cause harm to the scientific community and damage your own reputation.
* Contact the journal: You should contact the journal's editor or publisher as soon as possible to inform them of the error or misconduct. Explain the situation clearly and provide any evidence or documentation that supports your decision to correct or retract the article.
* Provide a clear reason for the correction or retraction: You should provide a clear and detailed explanation of why you are correcting or retracting the article. This can include minor errors, errors in the data, methodology, or analysis, as well as ethical or legal concerns.
* Include a correction or retraction notice: The journal will likely require you to submit a retraction notice to be published alongside the original article. The retraction notice should include the article title, authors, and publication details, as well as a clear and concise explanation of why the article is being corrected or retracted.
* Correct the record: If the retraction is due to errors in the data or analysis, you should work with the journal to correct the record by publishing a correction or erratum.
* Be transparent and accountable: Retracting an article can be a difficult and embarrassing process, but it is important to be transparent and accountable. You should take responsibility for the errors or misconduct and work to prevent similar issues in the future.
* Follow ethical guidelines: Retracting an article can have serious consequences for the scientific community, so it is important to follow ethical guidelines and best practices. You should be honest, transparent, and respectful throughout the retraction process.

*Retraction*

Retraction should be a last resort and should only be used in cases of serious errors or misconduct. If you are unsure about whether to retract an article, you should seek advice from the journal's editor, a colleague, or a professional association. By taking appropriate steps to retract flawed or unethical research, journals can help maintain the integrity of the scientific record and protect the trust of their readers and the broader scientific community.

Retractions can also be issued by the journals or Editorial Pontificia Universidad Javeriana. The following guidelines outline the steps that journals should take when retracting a published paper:

* Identify the reason for the retraction: Journals should clearly state the reason for the retraction in the retraction notice. Common reasons for retraction include scientific errors, plagiarism, data manipulation, and ethical violations.
* Notify the authors: The journal should contact the authors of the retracted paper to inform them of the decision to retract and the reason for the retraction. The authors should be given an opportunity to respond and provide any additional information that may be relevant.
* Publish a retraction notice: The journal should publish a retraction notice that clearly identifies the retracted paper, the reason for the retraction, and any relevant details. The notice should be prominently displayed in the journal and clearly linked to the original article.
* Provide a detailed explanation: The retraction notice should provide a detailed explanation of the reason for the retraction, including any specific errors or ethical concerns. The notice should also include a statement about the impact of the retraction on the scientific record and any potential implications for future research.
* Correct the literature: The retraction notice should include a statement about any corrections or clarifications that may be needed in other publications that cited the retracted paper. The journal should also work with other databases and indexes to ensure that the retraction is properly reflected in the literature.
* Follow established guidelines: The journals edited by Pontificia Universidad Javeriana follow established guidelines for retraction developed by the Committee on Publication Ethics (COPE). These guidelines provide a framework for handling retractions and ensuring transparency and consistency in the process.

Consider sanctions: In cases where there is evidence of misconduct, the journal may need to consider imposing sanctions on the authors, such as banning them from submitting future papers to the journal or reporting the misconduct to relevant institutions.

#### Double-Blind Peer-Review Process

Our scholarly journals follow double-blind peer review policy to ensure fairness and impartiality of the review process. This policy ensures that both the reviewers and authors remain anonymous to each other throughout, thus reducing the potential for bias and promoting objectivity:

*Reviewer Anonymity:* Reviewers are not provided with the name, affiliation, or any identifying information about the authors of the article being reviewed. This ensures that the review is based solely on the quality and merit of the research presented in the article.

*Author Anonymity:* Authors are not provided with the name or affiliation of the reviewers. This ensures that the authors' opinions of the reviewers do not influence the submission and review process.

*Reviewer Selection:* Reviewers are selected based on their expertise and experience in the field of the article being reviewed. Our journal's editorial board selects reviewers who have no conflict of interest with the authors of the article.

*Review Criteria:* Reviewers are asked to evaluate the article based on the following criteria:

* Originality of the research
* Significance and relevance of the research to the field
* Clarity and coherence of the argument and presentation
* Methodological rigor and accuracy of the data presented
* Appropriate use of references and citations

*Reviewer Feedback:* Reviewers are asked to provide constructive feedback to the authors on how to improve the quality and clarity of their research. Reviewers are encouraged to be critical but constructive in their feedback and to provide specific suggestions for improvement.

*Editorial Decision:* The editor-in-chief makes the final decision on whether to accept or reject the article for publication. The editor takes into account the reviewers' comments and feedback, as well as the article's originality, significance and quality.

*Reviewer Confidentiality:* All reviewers are required to maintain the confidentiality of the review process and not to share the content of the article or their review with anyone else without the editor's permission.

#### Open Access and Self Archiving Policy

This scientific and scholarly journals system works under a Diamond Open Access model. This means, it does not require any publication, processing, access or subscription fees. Instead, it is entirely funded and supported by Pontificia Universidad Javeriana.

Postprint and self archiving are allowed (to post the published version anywhere) with no embargo under the following conditions:

* Published source must be acknowledged
* Must link to the original work
* Must indicate if changes have been made

The Publisher SHERPA/RoMEO license can be accessed in the following link: <https://beta.sherpa.ac.uk/publisher/1759>

Editorial Pontificia Universidad Javeriana is part of the Public Knowledge Project Preservation

Network (PKP PN) to deposit and preserve all contents published by using the LOCKSS program.

#### Advertising and Marketing

Editorial Pontificia Universidad Javeriana does not regularly invest in advertising for the Scientifia and Scholarly Journal System. However, it ensures that any possible advertising piece does not interfere with the integrity and quality of the published contributions. Our advertising and marketing policy statement includes the following guidelines:

Advertisements do not compromise the quality or integrity of our publication or compromise our independence as a scholarly journal.

* Advertisements are not misleading or deceptive, and should comply with all applicable laws and regulations.
* Advertisements do not imply endorsement or sponsorship by our journals or our editorial teams.
* Advertisements are clearly identified as such and are visually distinguishable from editorial content.
* Advertisements are targeted to relevant and appropriate audiences and not interfere with the user experience of our readers.

We reserve the right to reject or remove any advertisement deemed inappropriate, misleading, or inconsistent with our values and mission.

We believe that responsible advertising and marketing practices can help promote the visibility and impact of our journals and increase the visibility and accessibility of the research we publish.

**35.** [**Readiness of the judicial sector for artificial intelligence in Latin America**](https://cetys.lat/wp-content/uploads/2021/09/compilado-eng.pdf)

INDEX PREFACE AI, Justice and Digital Transformation Policies in the Latin American Public Domain CASE ARGENTINA Assessment of the readiness of the judicial system for the adoption of artificial intelligence CASE CHILE Artificial Intelligence and The Judiciary. Chile and its Pending Challenges. CASE COLOMBIA Analytical and exploratory framework CASE MEXICO Readiness assessment for the adoption of Artificial Intelligence in judicial systems CASE URUGUAY Preparedness assessment for the adoption of Artificial Intelligence in judicial systems of the region

**Justice and Artificial Intelligence:**

Conceptual foundations of the research The development of Artificial Intelligence (AI), especially in its forms related to automatic learning or machine learning, has become one of the most relevant challenges faced by contemporary societies. Its impact on different sectors is being assessed in a context of skepticism, dystopia and optimistic narratives, which may even misrepresent the possibilities and limitations of this technology. While the most advanced AI has become a pervasive, widely discussed element in the media in fields such as recruiting and marketing, it is permanently going forward in areas such as healthcare, transport and education (Pombo et al, 2020). This AI progress in different types of activities has promoted in the last years an emerging, albeit growing, discussion in the judicial domain as well. This sector has some characteristics that make it particularly appealing for the implementation of AI. Most notably, the great volume of information and data generated in the administration of justice render it a particularly relevant space to implement AI techniques that make it possible to systematize, deduce, and create patterns and prediction in less time and with greater resource efficiency. The introduction and implementation of AI and its advanced expressions based on machine learning, and in particular its sub-field known as deep learning, pose varied challenges for the legal systems in any jurisdiction (Chen et al 2019). There are distinctive characteristics of the operation of the Judiciary and its regulatory and cultural environments that should be understood in the context of an increasingly more widespread discourse regarding the inclusion of machine learning techniques in the different Latin American judicial systems. Firstly, and at the most basic level, technology is helping inform, support and advise the people involved in the justice system in the form of supporting technology. Secondly, AI systems may replace functions and activities that used to be performed by humans, which would be called “substitution technologies”. Lastly, at a third level, technology may change the way judges work and provide very different forms of justice, acting here as a disruptive technology, especially when processes change significantly and the predictive analysis may reshape the allocation function. At the second and third levels is where the main questioning related to the impact of technology on judges’ roles and functions regarding the allocation function is raised (Sourdin, 2018). This diagnosis is also shared with another report drafted at the European Commission that we are going to deal with later on. The introduction of AI in the judicial domain may have a transformation impact, with risks and benefits for a better governance of the legal system. As for the positive aspects, the use of AI in this field may have an influence on the internal organization of these systems, as well as on the administration of justice for citizens of these countries, making it more affordable, accessible, transparent and agile. It may improve accountability and efficiency, and reduce the workload (COE, 2019). At the same time, there are risks associated with the misuse of these systems based on the lack of knowledge of its operation as well as the lack of responsibility of decision-makers, designers of the technology

architecture, as there are problems inherent to the design and operation of these systems that revolve around accountability, equity, access, transparency and interpretability, many of which have been considered in practice and in the recent literature as ethical problems. These problems go beyond the use of AI systems in judicial realms and are more general for their implementation in different sectors. However, the criticality in a legal structure in democratic systems and for the full exercise of human rights deserves an even more careful consideration of the inclusion of AI systems in justice. As there is not a unique approach on AI as a tool to provide solutions for the challenges faced by the Judiciary in different countries of the region, nor is there evidence that its use in this sector is questioned, while there is no universal judicial model, this work aims at understanding the direction of discussions and practices, policies and regulations, many of them incipient, about the implementation of AI systems in the Judiciaries in the region. This work seeks to complement international discussions as well as bridge existing gaps about the region by means of empirical evidence on the state of the issue in AI systems in legal structures, considering the political, institutional and digital technologies adoption design. To that end, this work develops an analytical framework that makes it possible to map the national approach to this topic, taking into account the national ecosystem and the international macro level; the meso level, adjusted to the dynamics inherent to the judicial system as organizations; and the micro level, related to concrete players and institutional entrepreneurs in this subject matter. Given the current context of uncertainty, skepticism, magnified expectations and economic interests of providers and a flourishing industry, this study points at filling the knowledge void related to the impact of AI and its meaning for the judicial experience in the Latin American context. This framework takes six dimensions that consider the reference indices, regulations and the stra-tegy related to AI; the governance of the judicial updating processes; the diagnosis and the capacities to adopt AI; the existing conditions for the deployment of AI in the legal domain; and the sense of opportunity that AI entails for the judicial domain in each country. These six dimensions are addressed with 50 indicators, aspects which will be developed in the second part of this report.

#### History of AI updating, digitalization and inclusion, and links with the judicial sector

Digital technologies based on the internet are increasingly perceived as an instrument to make justice go forward. When properly implemented, they can enable access, reduce differences and serve a wider audience. Technology makes it possible to provide services overtaking traditional methods and improving the experience of the population when they resort to the Judiciary. Accessibility, transparency and efficiency are principles that may be derived from a conscientious and responsible implementation of technology to serve justice (COE, 2019). A balanced implementation of technology may help achieve a people-centered justice. In this sense, the Sustainable Development Goals (SDG), in its item 16.6, have explicitly addressed the challenge of “creating efficient and transparent institutions at all accountable levels”. Accountability is still a key priority to administer and impart justice, and AI implementations promise to achieve it by getting rid of the inherent hu man bias. However, many algorithmic techniques used for deep learning are based on “black box” approaches, which do not allow for accountability. In civil courts, it is unimaginable to meet the objective of the SDG 16.6 of failing to solve the delays affecting all the cases (even the simplest ones in most contexts) and to move forward towards effective solutions –following the Rule of Law Index (World Justice Project 2019). Based on the analysis of the data coming from the countries that are part of this study, postponements and delays are endemic in the bureaucratic organization in justice systems. This problem has grown and become more evident in the context of the COVID-19 pandemic. As a general purpose technology with diverse techniques and implementations that go from the automation of processes to tools emulating human intelligence, the scope of AI is ample, and the needs, as well as the institutional capacity and preparedness to include these technology sets, also depend on the existence and availability of other technical conditions: network connectivity, computers and information systems, digitalized documents and the like, to name a few. Now then, by and large, AI and Information and Communication Technologies (ICTs) themselves do not explain why and how change takes place in public service organizations (OECD, 2019). Thus, the concept of digital transformation emerges as a necessary condition to assess the capacity and preferences of judicial institutions to incorporate digitalization and its processes. Even though the speed and broadness of changes in public services through the adoption of digital technologies, including the administration of justice, have been in the agenda for almost two decades (Heeks, 2001), the current stage is evaluated from the concept of digital transformation and “disruptive technologies”, rather than “electronic government”, as was coined at the beginning of this century. Digital transformation entails institutional options that are neither lineal nor deterministic by the available technology (Filgueiras et al, 2019). Technologies and institutional processes interact to put together a complex framework of digital transformation (Fountain, 2001). In particular, the digital transformation based on AI, as well as the implementation of AI in the entire public sector, presents a special condition that shows the double relationship of the use of this technology for governments: on the one hand, the obligation of States to protect citizens against possible algorithmic damages strains with their temptation to increase their own efficiency. In other words, the balance between governing algorithms at the same time as it governs by algorithms (Kuziemski, Misuraca, 2020: 2). The introduction of AI is framed by several discussions about changes in the judicial environments, starting with the nature of the adjudicative function of judges, as well as the context in which lawyers, courthouses and other sectors are currently using technology (Sourdin, 2018). From the perspective of public policies of digital transformation and in line with Sourdin (2018), the environment, the types of provisions, terms, rates, the provision and processes of user identification affect the ins-titutional choices that opt for an interface or digital solution. Thus, the digital transformations of public services depend on institutional processes with a variety of preferences for digitalization. These preferences point to choices of digitalization processes of certain services, but not others (Filgueiras, Cireno, Palotti, 2019). The adoption of digital transformation policies in any organization of the public sector allows for new forms of mediation and institutionalization of practices that represent a window of opportunities to promote inclusion, efficiency and effectiveness for the provision of services. Nonetheless, this same literature points out that a digital transformation process takes place within institutional contexts where choices and decisions are made based on players’ different preferences. The challenge becomes evident if we add that these preferences are made up by their knowledge and understanding of the topic at issue. With such a complex and ubiquitous technology as AI, the need to provide greater understanding once it is incorporated in the information systems of the national judicial system becomes vital. The convergence of ethics as an AI governance instrument and approach has been a sustained element in the last five years, with the development of over 60 documents about ethical principles and considerations for the evolution of AI (Fjeld et al, 2019). However, the only specific instrument for the Judiciary was promoted by the European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe, and it was adopted by the end of 2018. It is the first European document that establishes the basic principles for the use of Artificial Intelligence in the national judicial systems, and it portrays both the fundamental values and the basic methodological requirements for the creation and development of algorithms.

Basically, these principles expressed in Europe have become some of the few guidelines that reflect on the connection between the AI and the Judiciary. These principles are: respect for human rights, that is to say, the design and implementation of AI should be compatible with the fundamental rights; the principle of non-discrimination, which prevents the development or intensification of discrimination processes towards individuals or groups; the principle of quality and safety, which refer to issues of data governance focusing on their quality and safety; the principle of “transparency, impartiality and equity”, which allows data processing to be accessible, auditable and understandable; and lastly, the users’ control, which makes it possible to avoid a prescriptive approach, and to ensure that users are informed players and have control over their decisions (COE, 2019). An additional aspect that the CEPEJ document highlights are the recommended uses of AI in judicial systems –with an European focus given its area of competence-, although they constitute a relevant basis for the development of future recommendations for the practices of the use of AI in Latin American judicial systems, as well as other regions. The report identifies four types of uses: those to be promoted, those that require considerable methodological precautions; those that should be developed following clear scientific guidelines; and, finally, those uses that should be considered with the utmost reservation. Each of them will be analyzed briefly below As part of the AI uses to be promoted are an improvement of the jurisprudence through natural language processing techniques and data visualization that enable the task of studying and looking for sources beyond searching with key words; additionally, techniques that make it possible to gain access to justice, such as chatbots, and the creation of new strategic tools related to the activity at courthouses, allowing for projections on budgets, personnel and key performance indicators. Those uses that call for considerable methodological precautions include applications that require help for the elaboration of scales and parameters in some civil litigation; support for alternative mechanisms to deal with disputes in civil cases; online mechanisms to deal with disputes; and the use of algorithms in criminal investigations to help in the identification of locations where criminal offenses take place. The third group of uses that should be considered is only additional scientific studies that include the profiling of judges and the anticipation in court decisions. Finally, according to the COE document, among the uses to take into account with the utmost caution are those that use algorithms in criminal cases to make profiles of individuals and regulatory quantifications. In the second case, these uses encompass the systematization and standardization of information regarding the set of decisions produced by colleagues, with the risk that a judge’s decisions may be predetermined or circumscribed by the set of data framed as precedents. This classification does not constitute a definitive guideline but a relevant mapping to establish the risk level in discussions and experimentations currently carried out in the region.

#### An exploratory analytical framework designed for the Latin American reality

There is a strong shortage of analytical instruments devoted to analyze the Artificial Intelligence capacities in state agencies, and a complete lack of instruments designed to enquire on institutional preferences about the subject matter. With regard to state capacities, the only analytical framework intended to assess the maturity regarding AI is the Government AI Readiness Index, developed by Oxford Insights and the International Development Research Centre (IDRC) in 2017 (it also has 2019 and 2020 editions). Now then, it is only an instrument intended for the analysis at a government level and, strictly speaking, its main basis is a kind of summary of other indices (such as the Networked Readiness Index or the UN e-Government Survey). Moreover, it is oriented towards the scoring of countries on a number scale. Listing countries in a ranking may be useful for an overall look about the state of the issue at a global level; however, when the intention is to delve deeper in the cases with the premise that through their details and contrasts we may learn more about them, their preferences and the problem under study itself, international indices do not offer the best heuristic alternative. Likewise, they tend to introduce in an underlying manner a lineal idea of the capacity maturity processes, in general, from the identification of a base line built around the development paths deployed by the main economies of the world (which have the resources to fund global research projects), biases that are later adopted as expected paths by the other countries (and which may be more or less controlled by researchers in each case). That is why we believe that is vital to enquire, in turn, about state preferences (for example, in terms of the institutional design of public policies or the adoption of different international cooperation standards), as well as the dynamics that explain their formation processes at a national level. This need is exacerbated when we enquire about phenomena such as AI use, where there is no universal

standard but an emerging “regime complex”. Regarding studies on state preference for AI use, however, the shortage pointed out for the capacity study only gets stronger. In its 2020 edition, the Government AI Readiness Index incorporated a sub-index which covers 34 countries, where the concept of “responsible use” of AI is explored. What countries understand by “responsible use” is already an initial problem of the preference processes. However, we deem it necessary to go much further to elaborate a framework where both aspects, preferences and state capacities, are addressed in a comprehensive way. That said, when the subject under study is not the preparedness to incorporate AI in governments but in the Judiciary, such shortage becomes a complete absence of available analytical frameworks. That is why the project started off the elaboration of an original analytical framework that would make it possible to explore Latin American cases, which was achieved from the joint work of all the researchers involved in the project during three intense discussion sessions held remotely between August and October 2020, given the pandemic scenario. The main challenge that came up during that first stage was how to design an analytical framework that makes it possible to address really diverse cases given their differences in institutional (for example, there are federal and unitary countries, which has consequences on the institutional design of the judicial system) and socio-historical terms (there are countries with varying degrees of economic development, and different patterns of international integration), and in terms of digital transformation trajectory. In order to overcome this obstacle, we have created a consistent framework with six dimensions and over 50 empirical indicators, ranging from the general to the individual approach; moreover, they make it possible to explore both the precedents that are to frame the AI rollout as well as the existing developments. The dimensions considered were the following: • General mapping of the case according to reference indices. This dimension intends to place the case in the universe, especially in terms of its trajectory regarding information development, digitalization of public services and openness for public data management. To that end, eight well-known international indices are considered as proxy, and they address different aspects regarding the problem at issue: the aforementioned Government AI Readiness Index and the e-Government Index of the UN, the Global Open Data Index designed by the Open Knowledge Foundation, the Network Readiness Index designed by the World Economic Forum, the index of the World Intellectual Property Organization, and the ICT Development Index drafted by the

International Telecommunication Union. • Even though this dimension is the first one addressed for the elaboration of each national investigation, in the reports we will find it as Annex I. This is so because none of the studies based on international indices addresses specifically the judicial sector, so reading them entails a complement for each national report. Now then, reading is recommended as it is useful as a fine introduction for the subject matter being analyzed.

Regulations and strategies about AI. This dimension is divided into three sub-dimensions: i) National regulation; ii) National strategies; and iii) International alignments in terms of regulatory standards. The regulations consider both specific rules about AI and those with a direct impact on AI governance, as well as rules about the Judiciary updating processes and the protection of personal data, and it explores the constitutional hierarchy of Human Rights agreements. The strategies consider both digital strategies as well as those specific for AI, whether they are related to the judicial domain or not. Finally, in terms of geopolitical alignments, the degree of adequacy to the European Data Protection Regulation, the endorsement of Convention 108 and the principles about AI use of the OECD are considered, as well as provisions related to AI in digital commerce agreements. • Governance of judicial updating processes, with emphasis on AI. This dimension is divided into three sub-dimensions: i) Governance; ii) Players; and iii) Political discussion. In terms of governance, on the one hand, the Judiciary institutional organization outline is explored and, on the other, its outline of data and statistical management. Regarding players, firstly, the focus is on public agencies that take part in the judicial updating processes at national, sub-national and federal levels (the last one, where it is applicable), agencies and networks that have dabbled in the development of AI in the judicial domain, coordination mechanisms of issues related to the digital agenda in the judicial domain (with emphasis on AI use) and formal mechanisms of strategic coordination about AI use at state level. Secondly, the focus is laid on alliances and networks among the players mentioned. Finally, as for the political discussion, the forums and environments of the state sector in general and the judicial one in particular are considered, where discussions about AI use in the State are taking place. • Diagnosis and development of capacities for the adoption of AI. This dimension is divided into three sub-dimensions: i) Digital judicial services; ii) Human talent; and iii) Institutional resources. In terms of Digital judicial services, the focus is on the developments regarding process digitalization, inter-operability and expert systems. Regarding Human Talent, the existing capacities of the judicial personnel regarding the use of complex digital tools, such as AI, are surveyed, as well as the training plans in the judicial career with official support that incorporate the topic. As for Institutional resources, the existence of a top state agency with AI leadership and appropriate budgetary resources, and a leading agency with an integral plan for the use of AI in the Judiciary is considered. • Existing conditions for the deployment of AI in the judicial domain. This dimension is divided into six sub-dimensions: i) Starting point; ii) Theoretical basis; iii) Axiological basis; iv) Success metrics; v) Risk matrix; and vi) Responsibilities. In terms of starting points, we refer to surveying the initiatives effectively existing to incorporate AI in the Judiciary (at any level), as well as its objective and need, that is to say, their justification and the objectives associated with them. Theoretical basis refers to the analysis of the concepts of “humanity” and “justice” used as a theoretical basis for concrete propositions around AI in the Judiciary. Meanwhile, the axiological basis examines the principles and values that AI is framed into, with emphasis on the considera-tions about ethics. The success metrics refer to metrics used in AI-based projects to assess results. The risk matrix refers to the risks considered by public policy makers as related to the use of AI in the Judiciary, as well as the criteria based on which the risks structures are built, and the safeguards designed to deal with the unwanted effects of the use of AI in the Judiciary. Finally, responsibilities refer to the provisions related to the allocation of responsibilities for AI use in the judicial domain. • AI as an opportunity for the judicial sector. This last dimension is divided into three sub-dimensions: i) Sense of urgency; ii) Applicability conditions; and iii) Reflection on the role of Justice. The sense of urgency is focused on the key challenges of the judicial system in the county that may be dealt with better through AI, and the reasons why the initiatives to incorporate AI may be an opportunity for the Judiciary, beyond the ones effectively existing. The aim of applicability conditions is to distinguish the areas of the Judiciary where it is and it is not advisable to move forward with the use of AI, as well as the issue of how appropriate the available AI techniques are for the problems to be solved in the Judiciary. Finally, the reflection on the role of Justice suggests exactly that: an analysis focused on the systemic and long-term impact that AI use may have on Justice for the role of judges.

Description of the five national reports The second stage of the research project, which took place between November 2020 and May 2021, was devoted to preparing the national reports. In April, two remote workshops were held to pool the preliminary versions. Each report received comments from at least two researchers from other research teams. The final versions of each report were prepared on such basis. As has been said, the same analytical framework was adopted as the basis for the five national reports. Now then, the general criterion established was for each research team to be able to adapt the framework wherever necessary, keeping the fundamental elements and justifying such methodological strategy, by virtue of the possibilities offered by each case (in terms of evidence of initiatives effectively existing and access to verification means). Likewise, the study admitted some leeway for styles diversity. As a result, each report is, at the same time, a unique case study and an input to identify interest contrasts and similarities, without being a comparative analysis. Below we will point out the most interesting findings of each report. Later, the main similarities and differences among the five will be mentioned.

The study about Argentina, created by Gonzalo Bustos Frati and Bruno Gorgone (CETyS), is centered on what they call “effective initiatives of institutional innovation associated with AI in the judicial domain in Argentina”. What they mean by “effective initiatives” are the lines of action of a courthouse or another judicial player of the public sector that have been translated into the use of at least one module based on an AI technique. Defined that way, there is evidence of at least four initiatives, all of them driven by judicial agencies of the Autonomous City of Buenos Aires (CABA): those explored by the Criminal and Misdemeanors Court No. 10, the Courthouse No. 13, and the Statistics Office of the Judiciary regarding intelligent anonymization mechanisms of legal documents; and the one driven by the program Prometea, which belongs to the Office of the Public Prosecutor (MPF, for its acronym in Spanish), which in the last years has sought to develop the judicial expert systems through the incorporation of machine learning tools, and it has managed to become a defined and sustained policy in the MPF. From a methodological point of view, it is worth mentioning that the interviews to players of the judicial ecosystem who are driving institutional change initiatives associated with AI use were a critical input of the study about Argentina, which adds an interesting ethnological element to it. We will highlight some key aspects of the study. The first one is that all the cases have been driven by judicial agencies of the sub-national domain, specifically in the CABA, a phenomenon that players themselves think stems from the fact that this district has more resources than the other jurisdictions because it is the wealthiest of the country in terms of GDP per capita. A second outstanding point of the Argentina study is that the institutional innovation processes analyzed are limited to pioneering trajectories with a limited scale and an early degree of development. The only exception is Prometea, a node that has managed to become a defined policy in the Office of the Public Prosecutor (MPF) and which, since 2017, has tried to develop the traditional judiciary expert systems through the incorporation of machine learning tools. Even so, Prometea has also had to resort to alliances, in this case, with the Artificial Intelligence Laboratory of the Law School (IALAB) of the University of Buenos Aires (UBA). On the other hand, the authors state that Argentina does not have defined public policies for the use of AI in justice at a national scale, or a structure of institutional incentives “whose goal is for judicial officials to think they are capable of and responsible for the use of AI”, for example, by means of awareness and training programs. In this sense, a vital third element of the report is that, in all the cases in which the implementation of AI techniques has been effective, players have resorted to the use of their own human, technical and material resources, and to the creation of formal and informal, intrastate or cross-sectional networks and/ or alliances with other players. The implementations of “intelligent anonymization” encouraged by two of the 31 criminal and misdemeanor courthouses in CABA (No.10 and No.13) were possible from an alliance with another node, whether it is from the judicial sector (Courthouse No. 13 and the Statistics Directorate of the Judicial Council of CABA), or from the priva-te sector with public interest purposes (Courthouse No. 10 and the Technology Cooperative Cambá). Prometea has also resorted to the forging of alliances with other nodes, such as the IALAB of the UBA. A fourth point in the study is the focus on the upward or downward nature of the institutional change dynamics. Thus, ordinary courts have promoted strategies that may be described from a bottomup dynamic, whereas initiatives such as Prometea describe an institutional change trajectory that is bottom-down by a high-hierarchy agency, in this case the Office of the Deputy Prosecutor of the MPF. The latter explore the figure of institutional entrepreneurs, for example a driver associated with the digital transformation. The fifth outstanding point of the Argentina report refers to the realm of ideas. In line with Isabelle Stengers (2008), who believes that “what ideas we use to think other ideas matters”, Bustos Frati and Gorgone wonder what the previous ideas and beliefs of local experts and authorities regarding justice and AI are, as these ideas may condition the way players think about (and adapt) the principles stated at global governance forums, or in documents of international organizations. On this basis, they identify that, when the time comes to include institutional innovation initiatives based on AI in a broader vision of their ideas about justice, there are at least two frequent characterizations in statements: increased justice and open justice: the former is more axiological, while the latter is technological.

The study about Chile was conducted by Carlos Amunátegui Perelló, Raúl Madrid Ramírez and Matías Aranguiz Villagrán, from the Pontifical Catholic University of Chile. The report points out that the country does not have a specific and complete legislation that regulates Artificial Intelligence, although there is a far-reaching and relevant regulation about digitalization and automation of judicial processes. In particular, the Electronic processing Act 20,886 from 2016 stands out, which established uniform standards for all the jurisdictions regarding the processing of judicial proceedings through an online platform called Virtual Judicial Office. In turn, the authors mention that the precedent that started the digital transformation process was Memorandum 91/2007 of the Supreme Court, which stipulated the “Rules for Courthouses that process with electronic folders”. In this regard, they stress that the work was influenced by delays in the processing of cases in different jurisdictions, and that it initially met some resistance by the officials of the Judiciary who had been in the institution the longest. Likewise, the study underscores the discussions started in the Future Commission of the Senate, made up by 50 experts in the subject matter from the civil society, to make up working groups on the topic. It also ponders that it is currently discussing a new Civil Proceedings Code, which incorporates important aspects about digitalization and automation, such as the new civil execution model, completely digital and with electronic auctions. However, it points out that the country faces significant challenges when it comes to take the next steps and automate proceedings. In terms of master documents, Chile has the Digital Agenda 2020, which sets a roadmap in the implementation of digital technologies, and it establishes 63 precise compliance indicators. So far, Chile has not developed any official strategy of Artificial Intelligence, but the authors highlight the fact that the Ministry of Science, Technology and Innovation are preparing one following a cross-sectional consultation process. The study also ponders other facts as signs that the country is trying to strengthen the digital transformation of justice, such as the incorporation of the license of Watson Explorer by the Judiciary in 2019, as well as the license of the Watson Knowledge Studio cloud in 2020 for data extraction from web pages (web scraping) to train machine learning methods. Another interesting example, which does not go along the same line of hiring proprietary software, is the agreement signed in 2020 by the Supreme Court and ten universities for the development of a project to update the jurisprudence database. The platform will make it possible, with the use of AI, to access information about Supreme Court sentences easily and quickly. The authors also focus on the reasons why they deem it necessary to develop a sense of urgency around the incorporation of AI in justice. From their perspective, the “demand for judicial services in Chile has not been met” as a result of two main factors: on the one hand, many citizens cannot afford the provision of judicial services because of their high costs, and, on the other hand, there are multiple conflicts that “cannot be tried at court within the current costs structure because taking them to court entails costs that are higher than the possible returns derived from the litigation”. They claim that Artificial Intelligence might help in both fronts. Finally, the report suggests a series of guidelines and principles to lead the prospective incorporation of AI in justice, among which we can highlight the “responsible transparency”, the traceability of decisions and enabling human intervention whenever necessary, to name a few.

The Colombia report was conducted by Daniel Castaño, researcher of the Externado University. The Colombian case is particularly interesting as it has already moved forward with concrete initiatives to incorporate AI in justice, even high-hierarchy agencies, which sets Colombia apart from Argentina. Likewise, it has established multiple regulatory instruments, albeit most of them non-binding, with regard to digital transformation. We identify the main points of the study. Firstly, the Artificial Intelligence systems that have been incorporated in judicial or procedural matters are identified and described: “Prometea” and “Pretoria” in the Constitutional Court, “Siarelis” in the Superintendence of Corporations in jurisdictional functions, and the “Fiscal

Watson” in the Attorney General’s Office. Pretoria is the result of two memorandums (2018 and 2020) signed by the Court and two Argentine universities: the UBA and the Rosario University. The experience of Prometea and the IALAB, already mentioned, has been its conceptual basis. In concrete terms, it is a tool designed to optimize the processing of great volumes of information that the Constitutional Court receives to enable the document management and strengthen the protection selection process, and to improve the quality of the statistical information in the country. Siarelis was implemented to be a guide for users to explore solutions for possible corporate litigations. Finally, Fiscal Watson was implemented to strengthen the Unique System of Criminal Information through the association of cases by means of unstructured databases. Secondly, the report shows that Colombia has a robust judicial framework in terms of digital transformation and social ownership of information and communication technologies (ICTs). Here we find multiple key instruments: i) National Development Plans and, in particular, the Strategic Plan for the Digital Transformation of the Judicial Branch 2021-2025, drafted by the Higher Judiciary Council, which proposes an electronic file model, a network justice system, information management mechanisms, change management methodologies and the use of technology for legal training and citizen services; ii) the Statutory Justice Act and Processing Codes, which encompass legal mandates that favor the implementation and use of ICTs for the management and processing of judicial processes to render access to justice easier and more agile, under the guidance of the Administrative Chamber of the Higher Judiciary Council; and iii) the documents drafted by the National Council of Economic and Social Policy (CONPES), which are non-binding but set important guidelines in terms of digital transformation for public agencies, and even different instruments with ethical tenets about AI. The Document CONPES 3975 from 2021, which provides a definition of AI, among other aspects, stands out in particular. Thirdly, following the hypothesis that the implementation of Artificial Intelligence in justice administration is not only an opportunity but a need, the study delves deeper into a series of recommendations in terms of public policy. We highlight the following: a) strengthen what the author calls “digital ownership” by justice officials; b) make up a multidisciplinary team with the responsibility for identifying the need they intend to address with the use of AI, and assess the current state in which the task is performed with regard to transparency, promptness, quality, coherence, consistency, efficiency, coordination, inclusion and participation; and c) the implementation of an ethical governance framework of Artificial Intelligence that involves privacy by design, users’ rights and good practices to handle personal information from the architecture. Regarding the last point, Castaño suggests that the algorithms implemented in the administration of justice should comply with the “X-by-design” approach (X= Legality, Privacy or Digital Ethics), according to which the legality, privacy and digital ethics should be part of the technologies, operations and architectures of digital goods and services in a holistic, comprehensive and creative manner. Finally, the study emphasizes that “justice is human and it should remain so”; therefore, it is not about replacing judges with prediction machines. Far from it, the study intends to identify the areas with the greatest potential for the implementation of AI. It focuses on easy, mechanical, repetitive tasks, whose error margin does not pose a risk for the exercise of fundamental rights. Namely, it refers to the automation of processing matters, such as document management, follow-up on notifications and the scope of rulings and other ordinances, to name a few.

The study about Mexico was carried out by the researching team of the Center for Economic Research and Teaching (CIDE, for its acronym in Spanish): María Solange Maqueo Ramírez, Jimena Moreno González, Olivia Andrea Mendoza Enríquez, César Rentería Marín, and the assistance of Claudio Andrés Ambroglini Gómez. The report stands out because of its comprehensive approach to the institutional complexity of the judicial domain in a country with a federal political regime. Thus, even though there are no cases of AI use in the Mexican justice, at none of its levels, the report lays the regulatory, doctrinal and institutional foundations to project its future integration in the judicial domain. Along this line, the study identifies different factors that hinder the best preparedness to incorporate AI. Mainly, the lack of a homogeneous model of the Judiciary along the federative entities, which is verified by a significant diversity in terms of size and design of courthouses, chambers and administrative; and the great asymmetry between the Judiciaries in their digital updating, especially regarding the production, storage and processing of data that are relevant for decision-making in the administration of justice. The differences are even verified in the ways each jurisdiction addresses the process of digital transformation, as is the case of the digital signature. The federal Judiciary has digital signatures (FIREL), but the state Judiciaries have taken different steps; some of them have developed their own digital signatures, whereas others are adopting the FIREL or similar as their standard. In turn, in many cases, the functionality of the electronic file management systems available locally is limited to capturing and consulting digital documents. A third contrast has to do with the availability of an administrative unit of legal statistics that is in charge of the production and processing of data and metadata of electronic files. Indeed, in this regard, the authors consider that the Global System of File Follow-Up of the Federal Judiciary constitutes a best practiceFrom the regulatory standpoint, even though Mexico does not have specific legislation about AI, let alone a legal framework that addresses the current administrative and civil responsibilities derived from AI use specifically, the report considers that the present constitutional and legal order encompasses enough elements to face, in a general way, the damages that AI systems may cause to people. In particular, with regard to the regulation about legal data management, the document identifies a trend in the last five years, since the passing of the general Transparency and Access to Public Information Act from 2016, to make several changes. Additionally, the mapping of multiple specific documents around this subject matter, issued by different players of the Mexican judicial system, stands out. Regarding strategies, the study points out that there is no national strategy that leads players of the digital ecosystem for an action plan or that gives guidelines for the development of AI in the public sector. In the national digital policy, outlined by the National Development Plan and its respective programs, no possible advancement is explicitly projected that are related to AI. Now then, Mexico has had an AI Strategy since 2018, but the authors stress that it was published in the last days of the Enrique Peña Nieto administration, so its implementation has been rather limited, while there is no evidence that the current government of Andrés Manuel López Obrador intends to follow up on it. Another outstanding aspect of the study is the identification of international sources of the Mexican Law on the subject. The most prominent ones are the alignment with the European Union (data protection in line with the standards of the European Regulation), the Council of Europe

(through the endorsement of Convention 108 and the so-called Convention 108 Plus) and the Organization for Economic Co-operation and Development (OECD) (through the AI Principles), among other forums. Likewise, its participation in the Digital 9 Consortium (together with Uruguay) and the Ibero-American Data Protection Network (RIPD) where Mexico has played a central role in the regional coordination is highlighted. Finally, a valuable aspect is the notion that the context of the COVID-19 pandemic has stirred up the interest of lawmakers to create regulatory frameworks that refer specifically to this topic, which is expressed in the existence of four legislative initiatives, which are currently awaiting a decision. On the other hand, even though some precedents have been identified in terms of legal expert systems, none of them are currently in place.

Finally, the study about Uruguay was conducted by Julio Lens, Sandra Segredo and Fernando Vargas, from the Catholic University of Uruguay Dámaso Antonio Larrañaga. The pivotal point of this report is the proposition that Uruguay seems to be undergoing a “twospeed technology reception or progress”. This is so because there are two different scenarios depending on the state environment. Thus, government agencies appear to be “reasonably receptive to new technologies and moderately agile in their detection and incorporation”, while in other state areas, especially in the Justice, such progress is not taking place, or else “when it does, it is in a particularly cautious manner, significantly slower”. Indeed, in terms of informational development and digital transformation of the Executive Branch, Uruguay is one of the countries with greater trajectory at a regional level. The study explores the participation of the country in Digital 9 (Mexico is also part of the forum) and its realization in the Digital Nations Charter, as well as the efforts of the Agency for the Development of the Electronic Management Government and the Information and Knowledge Society” (AGESIC). It also addresses the digital plans deployed since 2006, renewed today around the Uruguay Digital Agenda, which have been translated into measures such as the implementation of the plans Ceibal and Ibirapitá, a unique integrating portal that enables online state processing, the digital file in the Public Administration, electronic communication and notifications, digital medical histories and the use of blockchain in public records, among other development (Aoudief, Ast, Deffains, 2021). At the same time, the study seeks to describe the institutional precedents in the judicial domain and, based on the above, it is a descriptive exercise that yields clearly asymmetrical results. For example, at a Judiciary level, there is not any comprehensive electronic judicial file that makes it possible to implement judicial processes in a totally electronic format. In fact, here the precedents that affect technology governance in the judicial domain are rather landmarks driven by the State as a whole, such as the Public Information Access Act (2008) or the electronic file (2009), which “created the framework necessary to install a new model of management for courthouses and supporting offices”. The study makes an effort to identify and map the digital capacities developed in the judicial field, such as the National Jurisprudence Database (in place since 2008) or the Multisubject Courthouse Management System (an IT application to follow up on processing at courthouses and courtrooms, as well as the systematization, preservation and consultation of all the proceedings). Now then, the most suitable element singled out by the study is that, recently, the Jurisprudence Department of the Judiciary started a pilot trial with the Engineering School of the University of the Republic for the anonymization of court decisions. So far, there has been no definitive report about the results produced by the enterprise, but it is an encouraging initiative. Meanwhile, from the regulatory standpoint, the report makes it clear that there is no national regulation concerning the use of Artificial Intelligence, or for government matters, and even less an explicit regulatory reference about its implementation for the processes carried out by the Judiciary. There is, however, a National Strategy of Artificial Intelligence for the Digital Government, published in 2019, which lays down the principles for the development, use and implementation of Artificial Intelligence systems in the public sector in the country. The authors also highlight, as a “guiding document”, the text “Questions for the assessment of the algorithmic impact study”, prepared by the Information Technologies Area of the AGESIC in 2020, as it is a “guide that applies to the system projects that use automated learning for decision-making”. **Mapping of similarities and differences between national cases**

mework has allowed us to identify some common aspects and certain contrasts that are worth mentioning (see Figure 1) to map the state of the discussion, and of the preparedness in terms of ideas and imaginary scenarios, (pre) concepts, policies and legislation on this subject. Figure 1 – Summary of national reports National digital strategies Argentina, Chile, Mexico and Uruguay have government strategies. Colombia has a digital strategy for the Judiciary. AI government strategies Argentina, Chile, Colombia, Mexico and Uruguay. With different degrees of implementation and continuity in the policies outlined. AI strategies in Justice None of the cases analyzed, even though in Colombia the digital strategy for the Judiciary mentions data analytics and provides a definition of AI. Specific regulations for AI None of the cases analyzed, even though Chile has a specific provision in the DEPA; in Colombia there are doctrinal non-binding documents. Effective initiatives about AI in Justice Argentina and Colombia Use of expert systems in Justice Argentina, Colombia and Mexico (they are no longer in place in Mexico) Institutional entrepreneurs and upward institutional change dynamics through the use of AI Argentina Downward institutional change dynamics through the use of AI Argentina and Colombia The first thing to point out is that, in all the cases, a kind of digital transformation process of twospeed public agencies is verified, where the dynamism observed at government level is not verified in the Judiciaries. This is so because of different factors that each national report addresses differently, but, by and large, they agree on the opacity of the judicial practice, the lower citizen control and judges’ reluctance to change. Strategic alliances for the implementation of AI in the Judiciary Argentina (all the alliances are informal, except for those generated by the IALAB) and Colombia (formal agreements with Argentine universities). The Courts of Uruguay and Chile have reached agreements with universities that may translate into AI-based initiatives. Speeding up digitalization processes after the pandemic In all the cases. In fact, in Colombia the digital strategy for the Judiciary was drafted within the framework of the pandemic. Potential applicability of AI in Justice All the reports agree on the fact that the automation of some mechanical and routine tasks, where it is not necessary to interpret a complex context, is an opportunity, as it would make it possible to take better advantage of the resources of the judicial system. With regard to the evidence of effectively existing initiatives about the use of AI, it is only found in the cases of Argentina and Colombia. In the case of Argentina, they are initiatives driven by judicial bodies at sub-national level, whereas in Colombia we identify initiatives in the highest-hierarchy judicial agencies at national level, such as the Constitutional Court and the Attorney General’s Office. A key similarity, however, is the fact that initiatives in both countries have been possible as a result of the convergence or alliance between different players. That is to say, cross-sectional alliances appear as a vector that favors the digital transformation of justice based on AI use. That has translated into initiatives for the effective use of AI both in Argentina and in Colombia, but it can also be verified that the two most interesting initiatives around the use of AI in the Chilean and Uruguayan Justice are related to agreements reached with universities. In the case of Argentina, the role of institutional entrepreneurs, who drive processes of upward institutional change of their own accord, stands out. Also in institutional terms, which have necessarily been manifested in AI governance in the judicial domain, the fact that the two countries with federal regimes (Mexico and Argentina) showcase a special complexity is highlighted; such comple-xity is brought about by the co-existence of local and federal courthouses. Now then, if the report about Mexico keeps the focus at the federal level, the report about Argentina is forced to explore the sub-national dynamics, as it is there where we see the greatest dynamism in that subject matter. In the other three countries, we find unitary regimes, which favor the unification of standards and processes, even when, naturally, this does not entail a preference in favor of incorporating AI. On the other hand, none of the countries analyzed has a specific regulation on AI, or national strategies for AI in the judicial area. The most widespread practice is the existence of national digital strategies, albeit restricted to the government domain. The only country that has incorporated a specific digital strategy for the judicial realm is Colombia, in 2020, from the challenges brought about by the pandemic. Even though there are documents as national AI strategies in Argentina, Mexico and Colombia, none of them addresses aspects related to the judicial domain specifically. Additionally, there is only evidence that it is implemented in Colombia. Indeed, both in Argentina and in Mexico the strategy was published in the last days of the governments that promoted their processes of cross-sectional consultation and drafting (Mauricio Macri and Enrique Peña Nieto, respectively), so their implementation has not been materialized in the administrations of the new national governments yet, and they serve as reference documents rather than current public policies (Aguerre, 2020). As for national digital strategies, they can be found in the five countries under analysis, although some of them have greater precedents in the subjects (such as Uruguay, which has submitted several strategies since 2006). Another common point in some of the works is the idea that the use of AI in the judicial domain is an opportunity (and not just a risk) given the expectation that the automation of some tasks and areas (different from those where AI is not applicable) might make it possible to direct the human talent to the critical aspects of the judicial system. Here, the reports of Argentina, Mexico, Colombia and Uruguay concur. Moreover, there is some consensus in that the most complex tasks, like the ones that entail the interpretation of a context or the approach to a case for which there is no clear legislation, should not be automated, in line with the uses recommended by the CEPEJ of the Council of Europe (COE, 2019). Finally, there is agreement that human beings should always be responsible for the decisions made; therefore, they suggest that judges will see their roles modified but not replaced. In turn, a significant similarity is that all the reports see the Covid-19 pandemic as a contextual opportunity, as it has created a sense of urgency among political decision-makers. Indeed, all the reports agree on the fact that the processes of judicial digitalization have been expedited since the beginning of the pandemic, even when different journeys may be observed. By way of conclusion, it is relevant to point out that we find the most relevant progress in the discussion, objectives and aims regarding the implementation of this technology in the judicial domain in those fields where there is work based on dia-logue with other sectors and players. More important than the level of digitalization reached by a country or State is the capacity to integrate levels of discussion that consider the macro, meso and micro levels as a key dimension to coordinate the objectives of this “automation”, as well as the different perspectives associated. Based on the experience of this research, the Judiciary needs to address the inclusion of all the advanced AI technologies in their systems considering the broader narratives taking place around AI use in other sectors, in other regions and with other players, as the values that should be protected are fundamental safeguards of democratic societies.

**36.** [**National Artificial Intelligence Strategy**](https://cdn.www.gob.pe/uploads/document/file/1909267/National%20Artificial%20Intelligence%20Strategy%20-%20Peru.pdf)

#### Artificial Intelligence

Classic definition Due to the accelerated development of Artificial Intelligence (AI), its definition has been transformed to the point where some consider that AI is everything that has not yet been invented. In order to build its understanding, we will start by mentioning that the term was coined by Professor John McCarthy in 1955. For him, AI was "the science and engineering of making intelligent machines. " Around the same time, mathematician Alan Turing also proposed a test to evaluate whether a machine could impersonate a human in its interaction with a real human.

#### Definition for the Strategy

• In its early days, AI simulated human behavior through rules. However, more recent methods focus on making machines learn from data and their interaction with the outside world automatically. Currently, the OECD proposes the following definition: "An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy" . • To complement these two definitions, we would also like to mention that AI is influenced by and influences areas such as neuroscience, linguistics, robotics, among others.

#### Machine Learning (Automatic or machine learning)

• Set of Artificial Intelligence models trained to learn from data in order to predict outcomes or make decisions without being explicitly programmed to do so. • Some of the best known types of approaches are: supervised, unsupervised, reinforcement and association rule learning, etc. • Some of the best known algorithms are: artificial neural networks, decision trees, support vector machines, Bayesian networks, genetic algorithms, deep learning and others.

#### External Context

The data economy and Artificial Intelligence applications are being adopted around the world at such a speed that, for the last 5 years, several nations and multilateral organizations have taken a more conscious stance on the need to promote and govern this process through the design and implementation of National Plans, Strategies and Agendas. Evidence on international trends (\*): • In recent years many countries around the world have invested in Artificial Intelligence led by the

US and China. • Use cases in almost every industry are accelerating. • Since 2015, the number of AI-related jobs in the private sector has doubled • Academic offerings for specialization, Master's and Doctorate degrees have increased significantly. • AI publications are increasing significantly. • AI technology or consultancy offerings are growing at an accelerated pace • Investments in technological infrastructure to support increased adoption and use. • The number of PhD graduates with AI and Machine Learning has doubled. • Multilateral organizations and Nations are defining Ethical Frameworks and regulations specifically dealing with AI. • Private investment in AI projects is increasing. https://hai.stanford.edu/research/ai-index-2021 Photo: Pixabay External Context In our region, Brazil, Mexico, Colombia, Argentina, Chile, and Uruguay have already made progress in this area, with various documented initiatives. In all cases, the documents reveal expressions of interest in the direction that governments should take with respect to these phenomena, and many of them are expressed in public policies but with little empirical support or technical basis. However, the cases of Colombia and Brazil have made significant progress in the implementation of their plans, which include the creation of institutions or agencies specially designed to manage strategies, coordinate efforts, and channel public financing. In all cases there was a great effort of socialization, dissemination, and involvement with the different stakeholders.

The main fronts addressed for the adoption and use of IA were: • Closing specialized talent gaps.

• Ensure the provision of actionable data and operations infrastructure. • Promotion of adoptionoriented investments in sectors such as small and medium businnes and startups. • Definition of Ethical Frameworks and specific regulation.

#### External Analysis: Use Cases of Local Government

Improved citizen service • Several municipalities in Norway use chatbots to speed up customer service for simple requests from citizens. • The municipality of Copenhagen uses NLP technique to analyze and sort through incoming emails from citizens. • Decision support • The municipality of Espoo (Filand) developed a pilot project to plan the budget allocated for public health services.

• The municipality of Trondheim (Norway) developed an administrative case recommendation system with the aim of speeding up decision making given a similar request.

**External Context: Use Cases in Justice:**

Robot Judge • China has been employing artificial intelligence in the courtroom since 2017. A robot judge is used to hear specific cases such as trade disputes, e-commerce liability claims, and copyright infringement. To date, more than 3 million cases have been handled by a robot judge in

China. • System to predict future criminal recidivism • HART is a system developed by the Durham Police and Cambridge University researchers, using training data from 104,000 people who had been arrested for five years. • The system uses variables that focus on the suspect's crime history, as well as age, gender, and geographic area, to classify an offender as low, medium, or high risk of committing new serious crimes during the period of next two years. • HART was developed with the aim of reducing the number of people incarcerated, and of being susceptible to other forms of intervention that would be as or even more effective in reducing the risk of recidivism **External Context: Use Case in Citizen Security**

Use Cases in Public Transport • Within public transportation, include cameras and AI-based systems to be able to detect people fights, follow people, agitators, or suspicious behaviour or any activity that could be a danger to women. It could also detect areas that do not have much light. •

These systems could alert operators to react quickly. • AI System to predict crimes in specific places and times • PredPol, a Californian company that grew out of a project between UCLA and Los Angeles Police Department, defines "predictive surveillance as the practice of identifying the times and places where specific crimes are most likely to occur, and then patrolling those areas to prevent for those crimes to occur. • PredPol uses the historical data of a client's police department from a period of two to five years to train a machine learning algorithm, which is subsequently updated on a daily basis. Only three data points are used: crime type, location, and date/time. • According to PredPol, demographic, ethnic or socioeconomic information is never used. This eliminates the possibility of privacy or civil rights violations seen with other predictive or intelligence-based policing models.

#### External Context: Use Case in Education

AI adapted to the needs of each student • From school to college, AI could individualize the learning needs of each student. • The system could respond to the needs of the student, placing greater emphasis on certain topics, repeating things until the student masters it. • In general, helping the student to learn at her own pace, whatever this could be. • AI could give feedback to students and teachers • AI could give feedback to teachers and students about the results of the course itself. • Some AI systems are used to monitor student progress and alert teachers when there might be a problem with student performance. • These systems could give the support to the students who need it, and to the teacher, to find the areas where he can improve the instructions to the student so that he does not fail with the subject of the course.

#### External Analysis: Use Cases in Agriculture

• Digital production and agronomy Initiatives dedicated to crop monitoring through sensors and images from drones and satellites). • Crop planning and management Managing crops, irrigation and workers based on data and predictive analytics • Market access and financing Crop and cost forecasting, insurance, exchange applications and leftover recovery betterfoodventures • 50% water and energy savings • Reduction of losses on expensive fertilizers due to overwatering

#### Use Cases in Fisheries and Aquaculture

* Reduction of operating costs Reduce the cost of farm maintenance, optimization of inputs and resources. • Observe Technologies and Umitron Cell offer AI-based technologies for fish feeding frequency pattern detection • Managing fish health • Norway's Seafood Innovation Cluster developed an ML system to predict possible parasite outbreaks in fish farms **Use Cases in Forest Protection:**
* Forest resource management Forest monitoring through satellite images. ● SilviaTerra and 20tree.ai use deep learning algorithms to generate highly accurate forest maps and extract attributes such as tree species, size and diameter, etc. • Deforestation detection Prediction of illegal logging activities. ● Rainforest Connection and Outland Analytics use sensors and deep learning models to detect suspicious sounds such as chainsaws or heavy machinery.

**Use Cases in Telecommunications:**

* Equipment and Network Optimization Traffic monitoring to detect problems in the communications equipment or to self-manage the traffic ● Aria Networks and Avanseus use ML algorithms to detect traffic anomalies and anticipate potential network problems. • Improved customer service Implementation of automatic question and answer system that can reduce the response time to a network problem. ● Success.ai or Vodafone use chatbots systems to solve the most frequently asked questions regarding a network problem

**Use Cases for the Mining Industry:**

* Process improvement • Analysis of machinery operation to determine maximum performance. • Freeport-McMoran achieved a 5% increase in copper production by using ML to increase mill throughput at its Arizona mines. • Improved scanning • New methods for the analysis of images and other sources of information to find mineral deposits. • Datarock uses deep learning and computer vision to extract geological information from geospatial imagery. • Minerva Intelligence uses semantic web technologies to harmonize databases with geological information for easy analysis.

**Use Cases in Energy:**

* Energy demand forecasting • In the UK, National Grid and DeepMind developed an ML algorithm to predict load in the short term, reducing energy consumption by 10%. • Optimize power generation • GE Renewable Energy uses software that monitors and optimizes the operation of its wind turbines, increasing energy production by 20% **Ethics in AI:**
* Biases and discrimination • Data and algorithm evaluation • Multidisciplinary evaluation • Job automation • Difficulty of measurement - Informality • Promote digitization jobs • Promote training • Guidelines, worldwide guidelines and proposals • Identify negative stereotypes and usefulness of personalization (access to university, access to credits, medicine and others). •

Promote use in vulnerable population problems, for health and logistic applications • Constant updating

**Internal Analysis:**

In Peru, some progress has been observed in the development and adoption of Artificial Intelligence, mainly from large economic groups and a small sector of academia; while the public sector has made very few attempts . These use cases have been centralized mainly in Lima. These are some of the situations encountered : • The country's main economic groups are carrying out or have carried out AI projects . • There are AI projects in the private sector, most of those projects are Machine Learning, few in Deep Learning . • It has been observed that the number of publications in Scopus Indexed Journals with an AI component has quadrupled in the last 5 years in the country, mainly due to incentives provided by concytec, innovate and other public funds . • The main universities that publish on AI topics are in Lima, followed by Arequipa. • The areas of publication with some AI component are mainly Computer Science, Engineering and Medicine with 50% of the total number of publications . • Currently many Peruvian universities have master's programs in data science, in addition to master's and doctoral programs with AI courses .

**Main organizations using AI in Peru:**

• Main economic groups: Breca Group (Brescia family), Intercorp (Carlos Rodriguez Pastor), Belcorp (Eduardo Belmont). • Companies / Large Corporations: Banks, Telcos, Retail, Insurance, Mining / Fishing, Manufacturing, etc. • Startups: Quantum Talent, Emptor, Fitness Pass, Xertica,

Teckton Labs, Latin Fintech, Chazki, qAIra, SpaceAg, etc. • Consulting Firms: McKinsey, EY,

MS, DMC, Everis, Accenture, Globant • Technologies (and their partners): Google, Microsoft, AWS, IBM, etc. There is evidence that in some cases everyone uses Machine Learning, some NLP, little Deep Learning.

**Labor market situation in Peru Photo:**

Pixabay The recruitment of professionals with knowledge of AI (or related subjects) in different sectors: • Private Sector: Corporations, Companies, Startups, Consulting Firms, Technological Firms Salary ranges: 95% of the salaries are between S/. 2,000 - S/. 20,000 (min: S/. 2,000 1st quart: S/. 4,000 mean: S/. 6,000, median: S/. 7,500, 3rd quart: S/. 12,500) • Academia: As research professors in Engineering, Computer Science, Diplomas, Business Schools. • Government: As hired staff in public institutions like Government and Digital Transformation Secretariat, SUNAT,

MEF, Minedu, Comptroller General of the Republic, Judiciary and others. • Otros sectores: Multilaterales (BID, Banco Mundial, CAF), ONGs / Think tanks (Innovations for Poverty Action - IPA / JPAL, etc.)

**State of the local academy in terms of training of professionals with AI knowledge:**

Undergraduate Studies: • Traditional Computer/Systems/Computer Engineering, Computer

Science careers that have been updating their curricula. • Other related majors - Economics, Mathematics, Physics have begun to focus on foundational support or AI courses. Postgraduate studies: • PUCP - Doctorate Program in Computer Science • UNI - M.S. and Ph.D. programs in systems engineering with AI courses, M.S. in Computer Science. • URP - Masters in Data Science • UPC - Master in Data Science • UNSA - Master in Computer Science • UP - MBA in Business Analytics, Diploma in Analytics • UDEP - Diploma in Data Science (with Pompeu Fabra University) • UTEC - Masters in Computer Science with specializations, Diploma in Data Science for Business • Several institutions that teach Data Science / Analytics / AI

courses: DMC, UNI, Colectivo23, BetaHouse, Kurios, Instituto de Analítica Avanzada (Breca), Digital House, Crehana, etc

**Results of the Peruvian Public Sector AI Status Survey (January - 2021) Question:**

In which use cases do you apply AI or Data Science in your organization? • Prediction of criminal acts • Facial recognition for assistance taking • Use cases in agriculture such as identification of crop types using satellite imagery • Natural gas and electricity monitoring using AI and regulation automation with Big Data • AI In Social Programs • Control of users of basic services (electricity and water) • Tax applications • Risk management • Virtual assistants • Customs risk management **Applications with AI Components Developed During the COVID Pandemic in Peru:**

"Peru in your hands" developed in conjunction with private companies with expertise in mobile applications, artificial intelligence and data analytics: Team: (UNI), (UP), (UTEC), MIT, Stanford University, INSEAD Paris, Universitat Pompeu Fabra & Barcelona GSE & IPEG & CEPR, Tekton

Labs, Kambista, Sapia, Mr. Burns, Media Labs, AmigoCloud, Alicorp, the Korean Ministry of

Interior, the Andean Development Community and the Inter-American Development Bank.

Project "COVID: Dynamic Virus Control". Winner of the "Special Projects: Response to COVID19" contest organized by Concytec. Proposal: dynamic control of virus infection in the Peruvian population with the use of artificial intelligence. Two processes: First: detection of potential COVID-19 infectees. Second: pre-diagnosis of the virus. Intelligent Integrated System for recording, reporting, alerting, monitoring and assisting people symptomatic of COVID-19 (SIAMA)." "SIAMA", a system to inform, alert, monitor and assist COVID-19 symptomatic people in Latin America, when delivering data and information does so through its App or web platform, generating automated alerts. The innovative system has a voice assistant, which, by means of periodic questions, will be informing the nearest healthcare personnel of the infected inhabitant. What technologies do you use? SIAMA uses Deep Learning (AI technique) that integrates a chatbot based on natural language. It provides patients with a better order in the level of urgency, in addition to the integration of the voice assistant (Google home or Alexa), which allows the interaction of symptomatic people, through voice commands; Smart Security Office" platform to control Covid-19 Company: MDP Consulting Winning project of the InnovaCovid-19 Challenge, of the Innóvate Perú Program, will receive a co-financing of up to S/. 450,000 Manages and controls the access of employees and visitors to work centers, through identity validation, detection of mask use, body temperature measurement and rapid triage for the prevention of Covid19. Technology: Facial recognition, temperature sensor to monitor the worker's health status and detect if he/she is wearing a mask. It allows for periodic triage of personnel through a chatbot

SWOT Analysis: Strengths • The country has scientific, mathematical and engineering talent. • Large economic groups interested and investing in AI • Government interested and promoting through PCM, CONCYTEC, FONDECYT, Innóvate and others. • There is a network of contacts with universities and top companies. • Programs such as Beca Presidente • Peruvian universities researching these topics Opportunities • Online Training (MOOCs) • Growing labor demand in the private sector, public sector, and academia for these AI positions (including from major companies such as Amazon, Google, etc.). • Growing interest and funding from Multilateral organizations • Peruvian Diaspora in top places • More Peruvians studying and researching IA in the world • The needs of attention to SMEs Weaknesses • Lack of high-level Educational Opportunities • Very little development of advanced AI (Deep Learning) in the private sector. • Incipient government on IA issues • Data situation at the country level, and at the level of organizations (private, NGOs, etc.). • Soft skills, English, and talent research capabilities • Salary incentives for academic researchers in AI are low. Threats • Brain Drain: growing labor demand in the private sector from major Amazon, Google, etc. • Other countries more AI advanced than us • Political instability •

Few institutions in Peru promote AI development • External sector attracts "all" talent • Peruvian society may have some fears about the use of AI as its use is further promoted.

**Purpose of the National Strategy for Artificial Intelligence:**

For a country like Peru, the adoption of new technologies such as Artificial Intelligence represents a historically unprecedented opportunity and at the same time a great threat to be left behind in the global socio-economic development and to deepen our shortcomings. This historical singularity presents an enormous responsibility for all members of our Society, not only the State, but also for the Private Sector, the Academia and the Civil Society. Then it becomes imperative the will to govern this adoption process with a National Strategy for Artificial Intelligence (ENIA) that facilitates the conditions to take advantage of these opportunities and mitigate the risks derived from this process. It is clear that the functioning of Society has been digitized on almost all fronts, driving a new economy whose main asset is Data. Creating value with these assets is an imperative not only for the private sector, but also for the public sector, which must also meet the new demands of a digital citizenship. In this context, the most advanced nations have deployed efforts during the last 4 years to express their intentions to govern a process that turns AI into a tool for socio-economic development and that is not a consequence of chance. This has been the main purpose for which various manifestations have been proposed, ranging from Strategies, Plans and National Agendas that in their first stages contain diagnoses of the starting point and aspirations for the end of this decade, through public policy recommendations, strengthening of some sectors and prioritization of public-private initiatives to ensure the first steps towards a National Artificial Intelligence Strategy. The following National Artificial Intelligence Strategy (ENIA) is proposed for the period 2021-2026, which can be updated every 2 years according to new technological advances and the situation of the country and the world. Peru is recognized as a Latin American leader in research, development, innovation, deployment, use, adoption of AI, and in its ethical and responsible use in the production of public and private goods and services. These efforts aim to accelerate national development and promote digital inclusion while ensuring the reduction of social gaps. Implement the National AI Strategy, promoting the development of human talent, infrastructure and the production of goods and services based on AI for the benefit of a more inclusive and multicultural Peruvian society that takes advantage of Industry 4.0 technologies in favor of sustainable development

**Strategic objectives:**

#### E1 - Training and Talent Attraction

SO.1.1. To position Peru as a country that enhances its human talent at all educational levels for the research, development and uses of Artificial Intelligence in the country. SO.1.2. To Lead regional research, scientific publication and patent publication in AI in key sectors of the country. SO.1.3. To be an attractive country for AI research and development. SO.1.4. Reduce the gap of participation of women and minorities in AI training programs.

SO.1.1. To position Peru as a country that enhances its human talent in Artificial Intelligence at all educational levels for research, development and use of AI in the country. • A.1.1.6. Create master's and doctorate programs in AI / ML in universities in the country with the formation of CORE programs on Artificial Intelligence. • A.1.1.7. Create scholarship programs for Doctoral students from IA and ML programs. • A.1.1.8. Promote massive up-skilling and re-skilling programs in digital skills, and in ethical issues of data processing. • A.1.1.9. Create courses or diplomas for the training of talent in the installation, administration and infrastructure support of high-performance supercomputers. • A.1.1.10. Create courses in parallel computing, signal processing and other courses related to high performance computing within the undergraduate and graduate programs of computer science, ing. software, ing. systems, ing. Computer science and other related.

SO.2.1. To Lead regional research, scientific publication and patent publication in AI in key sectors of the country. • A.1.2.1. Increase the number of publications in IA / ML according to RENACYT standards (in international conferences). • A.1.2.2. Encourage the creation of patents resulting from competitive funds and / or alliances with the private sector. • A.1.2.3. Facilitate international AI conferences that are held in different regions of the country on the following topics: AI applications in agro-industry, health, mining, forestry protection, energy, fishing and government, and in other sectors that the National Center for Innovation and Artificial Intelligence recommends it. • A.1.2.4. Facilitate visiting professor programs between local and foreign universities (including professors who teach remotely at provincial universities)

SO.1.3. To be an attractive country for AI research and development. • A.1.3.1. Create programs to attract Peruvian or foreign talent with Doctorate degrees in AI/ML from the academy, private or public sector, implementing incentive programs; for example, tax incentives for the private sector, or research funds for academia. • A.1.3.2. Create programs to retain talent such as PhD in AI/ML graduated in Peru or abroad, in research centers and universities in the country. These policies could consider better salaries, education and training opportunities, job stability and appointments, infrastructure improvements and others. • A.1.3.3. Facilitate the National Center for Innovation and Artificial Intelligence to repatriate national or foreign talent for its research and development projects in AI. • A.1.3.4. Facilitate public Peruvian universities to enter into agreements with top universities in Europe or USA in their AI/ ML master and doctorate programs.

• A.1.3.5. Facilitate the processes of validation of professional Master's or Doctorate degrees carried out abroad

SO.1.4. Reduce the participation gap of women and minorities in IA CORE training programs. • A.1.4.1. Promote the largest number of female students in undergraduate and graduate CORE

Artificial Intelligence (AI) training programs at Peruvian universities. • A.1.4.2. Promote the decentralization of AI/ML master's and doctorate programs to different regions of the country, promoting in regions where there is great potential as part of the country's key sectors, such as agro-industry, cattle raising , mining, forestry protection, health, fishing. • A.1.4.3. Create programming teaching programs for the unemployed and adults through regional or municipal centers in the different regions of the country.

#### E2 - Economic model

SO.2.1. Leading the research and development of Artificial Intelligence at the regional level. SO.2.2. Promote in public bodies, the incorporation of artificial intelligence in their operation and services to citizens. SO.2.3. Promote the integration of AI in the value chain to promote business development in the country's key economic sectors. SO.2.4. Minimize the effect of job displacement due to the adoption of AI

SO.2.1. To lead the research and development of Artificial Intelligence at the regional level. • A.2.1.1. Increase research funds in AI through the different programs financed with public resources. • A.2.1.2. Improve the evaluation of research and innovation projects with public funds with AI components by convening more experts in the area. • A.2.1.3. The winners of projects financed with public resources must upload their data and codes (models) to the open data platform of the government of Peru to be shared. • A.2.1.4. Promote scientific dissemination initiatives in coordination with the different research centers and universities. • A.2.1.5. Monitor the index development and adoption of AI (\*) and act where necessary to be in the first place at the regional level.

SO.2.2. Promote in public bodies, the incorporation of artificial intelligence in their operation and services to citizens. • A.2.2.1 Develop online courses for public officials in the adoption, use and benefits of Artificial Intelligence. • A.2.2.2. Promote the use of technologies such as chatbots or virtual assistants in public administration, prioritizing libraries and open source software. • A.2.2.3. Through the National Center for Innovation and Artificial Intelligence, prioritize the development of use cases where Artificial Intelligence can generate concrete solutions such as those proposed in various investigations aligned to the United Nations 2030 sustainable development goals, such as the elimination of poverty, zero hunger, quality education, clean and accessible energy, clean water and sustainable cities, good health, better qualified jobs, the reduction of social gaps and others. • A.2.2.4 Promote citizen participation in IAckathon, Datathon or Hackathon events organized by public bodies to create new services or improvements in citizen care, or to solve some problems in our country such as the detection of corruption in the public sector. • A.2.2.5. Promote the use of open source technologies in the public sector for the development of use cases based on Artificial Intelligence.

SO.2.3. Promote the integration of AI in the value chain to promote business development in the country's key economic sectors. • A.2.3.1. Create public funds and increase those that currently exist, for startups and basic or applied research programs that have some AI component, prioritizing the applications and areas of knowledge that the National Center for Innovation and Artificial Intelligence considers key. • A.2.3.2. Promote collaboration between academia and industry (micro, small and medium enterprises), also integrating CITEs from all regions of the country. • A.2.3.3. Promote initiatives to generate local use cases based on Artificial Intelligence with open data from the private, public and academic sectors of the country. • A.2.3.4. Promote a labelling industry through financial incentives for new ventures as part of publicly funded startup financing programs. • A.2.3.5. Prioritize and promote the adoption of AI in strategic economic sectors for the country, where we have comparative and competitive advantages.

SO.2.3. Promote the integration of AI in the value chain to promote business development in the country's key economic sectors. • A.2.3.6. Incorporate in each project financed with public resources the use of the infrastructure of the existing local data centers in the country, or in cloud platforms of private providers but where their technological infrastructure is installed in the country. • A.2.3.7. Promote research on Natural Language Processing algorithms in Spanish and other native languages. • A.2.3.8. Promote incentives for the private sector to publish open data on the country's national open data platform. • A.2.3.9. Promote regulatory sandboxes for key economic sectors of the country as environments for business development (small, medium, large companies, as well as startups) with few regulations in use cases with AI components, promoting their development and innovation in the country.

SO 2.4. Minimize the effect of job displacement due to the adoption of AI. • A.2.4.1. Create an observatory for the labor market to obtain timely evidence on the effects of AI. • A.2.4.2. Promote policies for upskilling / reskilling of the population that will be affected by the adoption of AI

#### E3 - Technological infrastructure

SO.3.1. Improve local infrastructure for AI research and development. SO.3.2. Promote the integration of local infrastructure. SO.3.1. Improve local infrastructure for AI research and development • A.3.1.1. Consolidate the National Fiber Optic Backbone Network project, in addition to installing a high-speed fiber optic connection between research centers. • A.3.1.2.

Promote the Implementation of 5G technology by spreading its benefits. • A.3.1.3. Create a National Center for High Performance Computing for research from academia, public institutions and industry in different areas of knowledge. • A.3.1.4. Promote the public-private partnership for the installation of the infrastructure of data centers in the cloud in the country with services for the academy, public sector, private sector and the citizenship in general of the country. • A.3.1.5. The government will promote contracting for the public sector, academia, local small and medium businesses, cloud services but with a provider that has local infrastructure and at a low cost. The government could guarantee a minimum monthly demand to be able to incentivize the installation of local cloud infrastructure in the country. • A.3.1.6. Increase investment in educational infrastructure for the training of new talents in AI that includes implementation of IA / ML laboratories in the different universities of the country

SO.3.2. Promote the integration of local infrastructure. • A.3.2.1. Through the National Center for High Performance Computing to govern all high performance computing centers and those created in recent years with public funding. The supercomputers have to be connected with a high speed optical fiber. • A.3.2.2. Rationalize the use of public funds for the acquisition of infrastructure equipment for research by facilitating the use of high-performance computing centers or local data centers once implemented. • A.3.2.3. Train local talent in the installation and administration of high performance centers in different regions of the country. • A.3.2.4. Create incentives to promote the use of high performance computing centers and local data centers in the country in research and development programs with public fund

#### E4 - Data

SO.4.1. To be a regional leader in the publication of open data. SO.4.2. To be a regional leader in the publication of data on biodiversity, native languages and other minorities in the country

SO.4.1. To be a regional leader in the publication of open data. • A.4.1.1. Create a rewards or penalties program for public and/or private organizations to publish high-quality open data in an open, reusable format on the country's government's open data platform (datosabierto.gob.pe). • A.4.1.2. In the datosabiertos.gob.pe platform include a module for communities to store codes and AI models based on the datasets published. • A.4.1.3. Create APIs to extract information from the single platform gob.pe and datosabiertos.gob.pe. • A.4.1.4. Monitor existing open data barometers

(\*) and act if necessary to keep the country in the top positions at the regional level. • A.4.1.5. Promote the monetization and benefits of data produced by the country's private sector, and its publication on the national open data platform. • A.4.1.6. Create a communication channel so that the academic sector, private sector, public or civil society in general, can request the publication of new open data, propose the publication of their data on the platform, and provide feedback on the quality of the published data

SO.4.2. To be a regional leader in the publication of data on biodiversity, native languages and other minorities in the country. • A.4.2.1. Create a bank of native languages, sign language and images to guarantee access for all citizens to public services (web, office, etc.) using translation and subtitling. • A.4.2.2. Create a database contextualized to each region of the country, for example: agro-industry, information on natural disasters, climate change, information on local industries, socioeconomic information, etc. This database will be published on the national open data platform. • A.4.2.3. Promote the creation of a database of the biodiversity of our natural and cultural wealth, in addition to promoting the investigations of these databases. These databases will be published on the government's open data platform

#### E5 - Ethics

SO.5.1. To be a regional leader in the responsible use of data and AI algorithms

SO.5.1. To be a regional leader in the responsible use of data and AI algorithms. • A.5.1.1. Adapt the recommendations of the OECD "Principles on AI" (\*), which Peru has signed, to the national reality, and start their implementation prioritizing research and development that stimulates the innovation of a reliable and accessible AI, prioritizing the sectors that the country considers key. • A.5.1.2. Through the different regulatory bodies of public services and national superintendencies, evaluate an impact study on the less biased use of algorithms for the classification of people in the private sector. • A.5.1.3. Create an observatory to monitor, report on rankings indicators for the responsible use of AI such as the Oxford Insight Responsible AI ranking (\*\*) and others. • A.5.1.4.

Create a unit to monitor and promote the responsible and ethical use of AI in the country.

SO.5.1. To be a regional leader in the responsible use of data and AI algorithms. • A.5.1.5. Implement a platform to be a registry of AI algorithms used in public sector. In addition to algorithms, the datasets used in use cases will be included. • A.5.1.6. In regulatory sandboxes created as a controlled environment where AI-based ventures can be promoted, promoting the ethical and responsible use of AI. • A.5.1.7. In the public sector, in all cases of use of AI to classify people (to provide benefits, opportunities or sanctions to citizens), they must have a socioeconomic impact study to guarantee equity. • A.5.1.8. Promote Ethics courses in all undergraduate and graduate programs in Computer Science, Software Engineering, and in all programs that contain AI courses

E6 - Collaboration SO.6.1. Facilitate the exploitation of synergies between universities and research centers through inter-university cooperation, the private sector, public organizations and international organizations SO.6.1. Facilitate the exploitation of synergies between universities and research centers through inter-university cooperation, the private sector, public organizations and international organizations. • A.6.1.1. Research projects presented to public funds such as Fondecyt, Concytec and others, must include as evaluation criteria, the participation of at least two universities, one from Lima and the other from the provinces. • A.6.1.2. Promote collaboration with foreign universities, and include some criteria for the delivery of public funds to universities in CORE training programs, promoting the participation of prestigious universities in the US and

Europe in master's and doctorates training programs. • A.6.1.3. Through a National Center for Innovation and Artificial Intelligence to promote collaboration in training, research, development and innovation of AI between the public and private sectors, academia, and alliances with prestigious foreign institutions. • A.6.1.4. Create and maintain a public registry of public, private and academic entities with AI/ML capabilities and conducting AI/ML research and development in the country. • A.6.1.5. Promote an alliance between countries in the region for the research and development of artificial intelligence. For example, explore a Pacific Alliance for Artificial Intelligence.

#### NATIONAL CENTER FOR INNOVATION AND ARTIFICIAL INTELLIGENCE

Purpose It is a national center of excellence in artificial intelligence, in response to the need for human talent and with the aim of accelerating the development and adoption of AI, which will carry out research, development, socialization and innovation by adopting Artificial Intelligence, coordinating its activities with the national and international academic sector, the private and public sectors, considering the sectors where Peru has competitive advantages, whether due to biodiversity, key economic sectors or criteria that are considered important for the country; inserting the country into a global value chain and promoting its integration into a globalized and digital world. It will prioritize the use cases where Artificial Intelligence generates solutions to the country's objectives such as the elimination of poverty, zero hunger, justice, quality education, health, clean and accessible energy, clean water and sustainable cities, the reduction of social gaps and others. use cases Functions • Carry out Research in Artificial Intelligence, as own projects or in collaboration with the academic, public and private sectors. • Carry out and promote the training of talent and skills at the national level in the research, development and adoption of AI framed within the ENIA. • Recommend and prioritize areas and cases of use of AI in coordination with the public, private and academic sectors of the country. • Carry out and facilitate AI job postings.

• Sponsor national and international academic AI events.

#### NATIONAL HIGH PERFORMANCE COMPUTER CENTER

Purpose The National Center for High Performance Computing is the national center specialized in high performance computing (HPC) that will manage the country's super computers. This center will provide high-performance computing services to academia, the private sector, the public sector, and industry, and will also provide advanced training in high-performance computing, promoting their participation. The main players in this center will be universities and national research centers that will be connected to the centers through a national high-speed fiber optic network.

**37.** [**Guidelines for the use of artificial intelligence tools in the public sector**](https://minciencia.gob.cl/uploads/filer_public/ae/9a/ae9a7ce7-807b-4781-9ac3-9b253bfbe735/of_n711_2023_dis_lin_ia_minciencia.pdf)

FROM: MINISTER SECRETARY GENERAL OF THE PRESIDENCY MINISTER OF

SCIENCE, TECHNOLOGY, KNOWLEDGE AND INNOVATION TO: AS DISTRIBUTED

#### BACKGROUND

The swift advance of artificial intelligence has shown that in a very short period of time, it represents a pivotal transformer. In the face of this development, the Administration of the State must assume a proactive role in the management of this technology, understanding how it functions and utilizing it in the most beneficial way for society. This must always be done ethically, consciously, and responsibly, with an emphasis on people, and respecting the fundamental rights of those who work for the State. Empowering the State in the responsive use of artificial intelligence is fundamental, offering powerful reasons to enhance efficiency in public administration and to foster civic empowerment. Similarly, it is key to develop and strengthen national competitiveness at an international level, to encourage investment, innovation, and entrepreneurship in new technologies. This field, after all, involves issues that affect social, political, and economic dimensions. It is essential to build a common framework that encourages the use and development of technologies based on artificial intelligence within the different bodies of the Administration of the State, and at the same time, respect the rights of the persons in the face of the implementation of these types of tools. This framework will contribute, at the same time, to the goals of digital transformation of the State. As the Government of Chile has committed to the agenda of modernization and digital transformation for the public sector by the end of the year 2022, the Digital Government Division (initially and subsequently, "DGD") of the Ministry of the Interior and Public Security (henceforth, the Ministry) will be in charge of coordinating the implementation of the Digital Transformation National Strategy, aligned with the regulation of the Law No. 21.180 of Digital Transformation of the State, coordinating the implementation of the Interoperability and Data Management System, implementing the State's Interoperability Network, and in conjunction with the Civil Protection Procedures and the Digital Readiness of the State, creating public policy 'ChileCompra', designing policies of Advanced Technologies and the Electronic Signature 'e-ChileCompra' and the 'Technological Compass for the Public Sector' ('TIC System').

1. In October 2021, the Ministry of Science, Technology, Knowledge and Innovation

(hereinafter and interchangeably "MCTCI") published its National Intelligence Policy Artificial with the objective of inserting Chile into the global collaboration related to the development of tools based on this technology, having an ecosystem of research, development and innovation in the field that creates new capabilities in the productive, academic and state sectors, and that, following the principles transversal aspects of opportunity and responsibility, contribute to sustainable development and improve the quality of life for everyone. This instrument is currently in the process of updating.

1. In light of the above, it has been deemed necessary to recommend to the bodies of the State Administration the following guidelines regarding the use of tools of artificial intelligence in the public sector.
2. These guidelines will constitute a first step to arrive at a common framework, since provide general guidelines and recommendations to promote responsible, ethical, safe and transparent tools based on artificial intelligence systems in the public sector. Its monitoring will allow preliminary and detailed information about of the regulatory and coordination challenges at stake and, if necessary, determine What instruments are most efficient and effective to address them.

#### RECOMMENDED GUIDELINES

It is recommended to the bodies of the State Administration that, in the exercise of their functions, decide to develop, implement or use tools based on information systems. artificial intelligence, consider the following guidelines, through their respective

digital transformation coordinators: Yo. Artificial Intelligence focused on people to. Evaluate whether the use of artificial intelligence is the most appropriate technological solution both to meet citizen needs and good service and functioning of the Administration that they are trying to satisfy.

1. Ensure that people are not subject to arbitrary discrimination based on their race, color, ancestry, gender, sexual orientation, age, language, religion, political opinions, national, ethnic or social origin, economic or social condition of birth, disability or any other reason or reason, in accordance with the provisions by the Constitution or the laws, due to the use of tools of artificial intelligence.
2. Involve citizens and/or civil society, when appropriate, in the design and implementation of artificial intelligence tools, especially when they have the potential to affect their rights, through public consultations and other participation mechanisms provided for in national regulations.
3. Follow the guidelines of the Guide to the Ethical Formulation of Science Projects of Data, published by the DGD, for the implementation of these measures and others as well as the actions that derive from them, when applicable.

ii. Transparency and explainability to. Inform citizens through suitable and easily accessible mechanisms about of the artificial intelligence tools that are being used within their processes, whether in the provision of public services or in supporting decision-making. decisions of the State Administration body.

1. Communicate through suitable channels, easy to access and in clear language, if they use or they use artificial intelligence tools in their relationship with citizens, such as chatbots, virtual assistants or automated systems for customer service. public, among others.
2. Make available to citizens, through suitable channels that are easy to access and in clear language, information regarding the use of tools artificial intelligence to exercise its powers, especially in the case of decision-making processes that could affect fundamental rights of

people.

iii. Privacy and data use to. Guarantee that the processing of personal data, especially those of a sensitive, which is made by using artificial intelligence tools, complies with the Law No. 19,628 on the protection of private life or its modifications, in particular, to ensure that data processed for the development, training or use of artificial intelligence tools are used exclusively for the purposes authorized by their owners or by law.

1. Do not enter personal information, especially that of a sensitive nature, in generative artificial intelligence tools, when these have not been contracted by the State Administration body or developed by or for this. In this regard, special care must also be taken with the information confidential information of legal entities to which the Administration has access.
2. Periodically review the studies, recommendations and jurisprudence of the Council for Transparency and other competent authorities in matters of protection of personal data, which are relevant for the implementation of these measures and the actions that derive from them.
3. Comply with the provisions of Decree No. 7, of 2023, of the General Secretariat Ministry of the Presidency, which establishes the Technical Standard for Information Security and Cybersecurity in accordance with Law No. 21,180; and Decree No. 11, of 2023, of the Ministry

General Secretariat of the Presidency that establishes the Technical Standard of Quality and Operation of the electronic platforms that support procedures administrative bodies in the State Administration bodies and those relevant recommendations issued by the Ministry of the Interior and Security Public on cybersecurity.

iv. Other measures to. Ensure, within the needs and possibilities of each service, training of its officials and other public servants of its organization for the adequate adoption of tools based on information systems artificial intelligence.

b. Pay attention to the guidelines, suggestions and learning contained in the following documents:

Yo. Government study of data in the Ministry of Social Development and Family of the DGD and the Inter-American Development Bank.

ii. Guide "Allowed to innovate: How can we develop science projects of data to innovate in the public sector?" from the Government Laboratory. iii. Publication "Transformative institutions 2023: good practices of the Central Administration" of the DGD.

1. For the purposes of following up on the recommendations made, the Head of the

DGD, in the exercise of its functions, may directly officiate the bodies of the State Administration to request any data or information related to these guidelines. Likewise, the Director of Data of the DGD may request, via email, to the Digital Transformation Coordinators, any data or information related to this instrument.

1. The delivery of information from the Digital Transformation Coordinators in response to the data requested from the Director of Data of the DGD will be made at the email box Gobiernodatos~a digital.gob.cl.
2. When applicable, the Digital Transformation Coordinators will inform that another person from your institution has been designated to assume the duties of communication and follow-up described in this section. In that case, the Director of Data from the DGD must make all information requests to the new person that is designated.

11 The MCTCI, together with the DGD, within its scope of competence, may develop guidelines for the design and implementation of AI-based solutions in the public sector and facilitate compliance with the measures contained in this instrument.

12. Any change that may occur regarding the organizational structure of the Departments of the State mentioned in this instrument, will not affect in any way case the validity and validity of the expressed measures. Their roles defined in the present office will be assumed by their legal successors.

**38.** [**ChatGPT: How to use it in classes?**](https://www.uc.cl/noticias/chatgpt-como-usarlo-en-clases/)

Strongly debated, the truth is that this tool is here to stay. Specialists agree that it simplifies and optimizes different tasks and jobs, both in the most diverse disciplines and in everyday life. How to use it in classes and take advantage of its advantages? Here are some recommendations.

What are you? A useful tool or a threat? We asked ChatGPT himself: “I'm an artificial intelligence guy. "My main purpose is to process and generate text in response to user questions and requests," he tells us and adds that yes, it is useful," he replies.

“I am designed to be a useful tool. My main purpose is to help and provide information on a wide range of topics. I can assist in text generation, answer questions, offer suggestions and provide support on various tasks,” she adds.

However, ChatGPT emphasizes that “my usefulness depends on the context and the specific needs of each user. It is important to remember that I am a tool and that it is always advisable to use human judgment and consult other sources when necessary to make informed decisions.

ChatGPT – if we use version 3 – “in practice, and in more general terms, is an artificial intelligence chatbot (virtual operator) created by the artificial intelligence research laboratory Open AI, and based on a large language model "specially trained to establish conversations," says Gabriela Arriagada, researcher at the Institute of Applied Ethics and the Institute of Mathematical and Computational Engineering.

As specified in the document “A brief look at the current state of Artificial Intelligence” , prepared by professors Marcelo Arenas , Gabriela Arriagada , Marcelo Mendoza and Claudia Prieto , the

GPT Chat “was trained to answer questions and generate answers in natural language . One capacity that Chat GPT has is to incorporate a conversation into the text generative mechanism, placing emphasis on the feedback obtained in the interaction with the human.” Users can ask you questions and ask you to prepare organized information to solve requests, such as a math problem or an essay about a book.

**A threat?**

So, is ChatGPT a threat? Professor Gabriela Arriagada explains that the way it works today, in versions 3 and 4, does not present a specific threat. As a tool, it simplifies and optimizes a series of tasks that can facilitate the work of different disciplines.

“It seems to me that the arrival of a tool that helps us optimize information searches, that helps us structure lists, make summary documents, review codes, etc., is welcome, as long as we understand what it does, how "It does it, and how to translate its use into a complement to our work and not a supplement," says the professor.

However, it must be considered, as the aforementioned document warns: “ A limitation that Chat GPT has is that it requires retraining to incorporate updated information. This limits their ability to interact around current issues.” Although version 4, launched last March, has “a greater knowledge base to interact around current issues. It is also capable of handling longer contexts

(…) favoring the creation of content, extensive conversations and automatic document analysis.”

As Engineering professor Marcelo Mendoza warns, on the one hand “it is an opportunity, because it produces an enormous advance in language technologies and represents a significant advance in the state of the art of natural language processing.” But at the same time it is also a threat, "in the sense that there are many aspects of GPT Chat that have not been studied, such as, for example, its ability to create biased content or produce disinformation." **How to use ChatGPT in classes?**

Surprise was what the Institute of Astrophysics professor Thomas Puzia saw on the faces of his students when he asked them to use tools like ChatGPT for the Astrobiology course.

“I wanted to motivate them, for this tool to be a support to search for references, review titles of books and articles, to make summaries of a lot of literature. This was a tremendous advantage for them. It was noted that the understanding of the material improved. I also promoted the use of artificial intelligence to prepare slides for an exhibition, which also improved the quality of their presentations,” says the professor.

The academic took into account that the use of this technology did not have a “good reputation”, but he proposed that this was not copy and paste, but rather he explained to the students that this was to understand, verify, practice logic and self-criticism of knowledge. In fact, he says his performance improved and his world expanded.

"With these tools, students can dedicate much more time to things that are more important, such as understanding concepts, their causality with each other, among other things. Without a doubt it has been a tremendous incentive for learning. This is the new world and it has arrived,” he says.

**What does the UC propose?**

As Chantal Jouannet , director of the Teaching Development Center, explains, ChatGPT can have very interesting uses to assist us in the areas of teaching management and generation of learning experiences. On the one hand, she says, it can be used as an assistant to facilitate teaching tasks, such as designing questions for the class, defining performance indicators, designing evaluation instruments, and more.

This, always using teaching criteria to analyze the relevance of the chat responses. “On the other hand, it can be used to design specific active learning strategies and in evaluations. For example, we can ask students to analyze and criticize a text prepared by ChatGPT 3. The chat can be used as a counterpart in a debate and can even be used as an object of study when discussing its operation and the ethical implications of its use. , Explain.

As Professor Gabriela Arriagada adds: “The basic criterion is to be clear about the purpose of its use, what am I going to use it for? Using it as an assistant is usually a good motivation, so being transparent in its use is important. “A good practice is always to declare what has been used and for what.” And she specifies: “Normalizing the declaration of the use of generative tools is the best way to combat the risks of plagiarism, originality, or distrust and fear of the integration of technology as a support tool.”

In the midst of the debate over the integration of artificial intelligence in the educational world, last May the Ministry of Education launched the new “Teaching guide: how to use ChatGPT to enhance active learning” . This document was born as a way to anticipate and address the challenges of new technologies to take advantage of the opportunities that artificial intelligence applications, such as ChatGPT, present in education.

“Another relevant aspect is maintaining a critical stance towards technology ,” says Gabriela Arriagada. And she adds: “This means understanding that, although Chat GPT is a great advance in contrast to previous technologies, we cannot trust that everything it shows us is perfectly acceptable.”

Chantal Jouannet highlights that the most important thing is to recognize that these tools have great positive potential in university education, just like other technologies in other times. However, she emphasizes the importance of its responsible use: “It is essential that there be dialogue with the student body about the contributions and limits of its use, as well as making it transparent when the performance of an activity has been carried out with the support of ChatGPT,” she explains. To investigate this topic further, both for students and academics, the Center for Teaching Development, the Academic Directorate of Teaching and UC Libraries, jointly developed a site with plenty of material and guidelines for the use of artificial intelligence tools, such as ChatGPT at UC , which can be reached here.

**What is the future of ChatGPT?**

“This technology is already embedded in services that we use every day. It is already connected to knowledge bases, so it will serve as an assistant for news writing, veracity verification (automatic fact-checking) and programming assistant along with tools such as copilot,” comments Professor Marcelo Mendoza.

As he adds, “the next step in these technologies is the generation of multimodal content . Currently the generation is unimodal, that is, you either generate text (GPT Chat) or you generate images (Dall-E). The production of bimodal content - for example, text + images - will allow the production of content such as concept maps or other content that combines text and images.

“It is possible that the quality of the responses, interactions, and results that these models and their chatbot counterparts can give us will be much more refined. It would not be unreasonable to see better chatbots implemented for customer services or to help with organizational tasks,” says Gabriela Arriagada.

“A future with supporting technology like this can change the way we work – especially if this is adopted massively – freeing us from simpler tasks, but demanding more of us with respect to more elaborate tasks ,” says the researcher and concludes: “We should not, however, talk about a future only of advances, but also of regulations and standards that support these innovations.”