Policy on the use of Artificial Inteligence in the Administration of Justice

2024

Preliminary version.

May contain translation errors.

POLICY ON THE USE OF ARTIFICIAL INTELLIGENCE IN THE ADMINISTRATION OF JUSTICE

1.1 AIM

In recent years, various artificial intelligence tools have been developed and implemented, both in the public and private sectors, also including in the Administration of Justice itself. Thus, and in line with this reality, Royal Decree-Law 6/2023, of December 19, has introduced automated, proactive and assisted actions in its Chapter VII, Articles 57, 58 and 59, which derives from the possible use of artificial intelligence to support jurisdictional actions.

This use by the Courts and Tribunals, guarantors of effective judicial protection, must be carried out with full respect for the applicable regulations, constituted by the Artificial Intelligence Regulation, or when applicable, the regulations for the protection of personal data, without forgetting the procedural rules themselves, means that minimum criteria must be adopted, both by those who are going to promote this type of projects and by the users themselves.

Consequently, through this document, a series of minimum criteria are established in order to ensure responsible, legal and ethical use of Artificial Intelligence in the field of the Administration of Justice, including an Annex with definitions, as well as a breakdown of accepted and other prohibited uses.

1.2 AREA OF APPLICATION

This policy applies to any system or service that is fed by judicial data, considering these as any that are produced within the framework of a judicial procedure, from the presentation of the first document or complaint to the final judicial resolution, also including possible publications, or notifications of the files.

For the purposes of determining the corresponding responsibilities, which are detailed below, it will be necessary to differentiate the use of those artificial intelligence tools that may affect the jurisdictional activity itself, and, consequently, affect judicial independence, from those that do not affect.

Furthermore, this use of judicial data, if it involves the processing of personal data, entails processing of jurisdictional data, as it affects the jurisdictional activity of Courts and Tribunals.

1.3 RECIPIENTS

All employees of the Administration of Justice must respect the content of this policy, and its adoption must be specifically ratified by the General Council of the Judiciary (CGPJ), State Attorney General's Office (FGE), Autonomous Communities with competence in Justice and the Ministry of the Presidency, Justice and Parliamentary Relations (MPJRC).

It must also be respected by the personnel at the service of providers of artificial intelligence tools, as well as any other institutional actor, public or private, who has access to the information held by the administrations with powers and CGPJ or is found hosted in the systems intended for the Administration of Justice.

1.4 BASIC PRINCIPLES THAT GUIDE THIS USE POLICY

The CEPEJ ("European Commission for the Efficiency of Justice") in its plenary session on December 3-4, 2018, adopted the document called "European ethical charter on the use of Artificial Intelligence in judicial systems and their environment", which contemplates five basic principles that must be followed.

POLICY ON THE USE OF ARTIFICIAL INTELLIGENCE IN THE ADMINISTRATION OF JUSTICE

These principles are the following:

1. Principle of respect for fundamental rights: when artificial intelligence tools are used to resolve a dispute or to assist in judicial decision-making, effective judicial protection, judicial independence and a process based on equity between the parties must be guaranteed.

Consequently, the development of these tools must be carried out applying ethics and respect for human rights in their design.

Furthermore, AI should never replace human decision-making in crucial matters of the Administration of Justice. It is essential that AI systems are used as tools to assist legal professionals in their duties, but the ultimate responsibility for making legal decisions must rest on Judges and Magistrates. Their decision-making must be ensured independently.

2. Principle of non-discrimination: it implies that it must be guaranteed that discrimination does not occur. It assumes that both in the development and application of artificial intelligence there is no discrimination based on the use of especially sensitive information such as data regarding racial, ethnic, socioeconomic origin, political opinions, religious or philosophical beliefs, membership in unions, genetic data, biometric data, health data, related to sexual life or sexual orientation.

In this way, learning these systems must ensure that this discrimination does not occur.

3. Principle of quality and security: for its correct development, the experience of professionals in general must be taken advantage of, such as that of Judges, Magistrates, Prosecutors, as well as researchers and professors in other fields of both Law as well as Social Sciences or engineering itself, and thereby achieve a multidisciplinary approach, sharing them among all ethical safeguards.

It also implies the use of certified sources, which will not be modified until they have been used in the learning mechanism, the entire process being traceable so that no modification occurs.

Both the models and algorithms created will be stored and executed in secure environments, to guarantee the integrity of the system and its intangibility.

4. Principle of transparency, impartiality and fairness: so that the development of these systems is accessible, understandable and auditable. To achieve this, a balance must be achieved between intellectual property and compliance with these principles of transparency, impartiality, equity and fairness, so that access to the design of these systems is allowed, biases are not produced, and the interests of justice are prioritised.

The algorithms used in the Administration of Justice must be transparent and explainable. Citizens have the right to understand how decisions that affect their lives are made, especially when it comes to judicial processes. Measures should be implemented to ensure that automated decisions are understandable and can be examined by experts and interested parties. It is highly recommended to offer advertising and transparency through the official websites of CTEAJE members. Specifically, it is desirable to publish FAT (Fairness, Accuracy and Transparency) records or similar instruments, about the data used, members of the AI teams, services, algorithms, possible biases and applications that make use of artificial intelligence techniques.

POLICY ON THE USE OF ARTIFICIAL INTELLIGENCE IN THE ADMINISTRATION OF JUSTICE

The use of state-of-the-art AI services that, for reasons derived from the novelty or cutting-edge of their technology, are not capable of offering high explainability may be assessed, provided that they do not represent a detriment to the rights of individuals, or the own jurisdictional functions and are fully in accordance with the applicable regulatory regime.

5. Principles "under user control": means that, although an artificial intelligence system is used, it must be guaranteed that judicial decisions can be reviewed previously. Citizens who may be affected by the system must also be informed, in clear and understandable language, whether the results offered by the tool are binding or not, as well as their possible use during the judicial procedure, before or during it, and their right to object so that they can be heard directly by a judicial body.

In addition to these principles promoted by CEPEJ, the following must also be respected:

1. Principle of Equity and Universal Access: AI in the Administration of Justice must be used to guarantee equitable access to judicial systems, regardless of location, socioeconomic status or any other demographic characteristic. This involves developing solutions that remove barriers and ensure that all citizens have the same opportunity to assert their rights before the law.

2. Principle of Prevention of Bias and Discrimination: AI in the Administration of Justice must be designed to avoid bias and discrimination. It is essential to carry out periodic evaluations of the algorithms to identify and correct possible biases inherent in the training data or algorithms used. Additionally, safeguards must be put in place to protect people's rights and prevent the perpetuation of systemic injustices.

3. Privacy and Personal Data Protection Principle: Strict privacy and personal data protection measures must be established in the use of AI in the Administration of Justice, including the need to previously carry out data protection impact assessments. Systems must comply with existing data protection regulations and ensure that individuals' confidential information is handled securely and ethically.

4. Principle of Responsible Innovation and Continuous Evaluation: Responsible innovation should be encouraged in the development and implementation of AI in the Administration of Justice. This involves conducting regular assessments of the impact of technology on the justice system and being willing to make adjustments and improvements as necessary to ensure its effectiveness and fairness.

5. Training Principle: It is essential to provide adequate training to legal professionals and other actors involved in the use of AI in the Administration of Justice. Understanding the ethical and practical principles related to AI is essential to ensure its responsible and effective use.

6. Co-governance Principle: Collaboration must be encouraged, sharing and exchanging knowledge, as well as the AI-based systems themselves between different areas of the organisation, or with the rest of the administrations with powers in matters of Justice, the CGPJ, and the FGE or with other institutions, in such a way that the desired innovative development is promoted in line with an ethical implementation of artificial intelligence.

1.5 RULES FOR THE USE OF ARTIFICIAL INTELLIGENCE SYSTEMS

The use of AI in the Administration of Justice must at all times comply with the different regulations, both

POLICY ON THE USE OF ARTIFICIAL INTELLIGENCE IN THE ADMINISTRATION OF JUSTICE

European and national, that apply to it.

In general, always keep in mind that:

1. The human review of everything generated whenever it directly or indirectly affects the rights of the users of the Public Justice Service or the judicial or jurisdictional activity itself.

2. The sovereignty of data and information must always be protected and respected. Neither the source nor the generated data/information should ever be accessible to (unauthorised) third parties.

3. Check that there is no bias in the tools.

4. All results obtained from generative AI must be identified as generated by an GenAI.

5. AI applications that have not been authorised should not be used.

6. Unapproved application programming interfaces (APIs), add-ons, connectors, or software related to AI systems, should not be installed.

7. The results of AI applications should be reviewed to ensure that they meet the organisation's standards for principles of fairness, ethics, and appropriateness.

8. Any results that discriminate against people based on race, colour, religion, sex, national origin, age, disability, marital status, political affiliation, or sexual orientation, should not be used.

9. AI applications to create textual, audio or visual content for the purpose of committing fraud or to misrepresent the identity of a person or the commission of any other unlawful act, should not be used.

1.6 AI RESPONSIBILITIES AND CONTROL

The different responsibilities that AI entails in the Administration of Justice must be determined. Thus, one can differentiate between:

Regarding the **use of AI**: the person responsible for its use will be the court, tribunal, judicial office or prosecutor's office, and within each of them, the person who makes effective use of it. Likewise, in the services offered to citizens or justice professionals, it will be the user themselves.

In relation to the **development and implementation of AI systems**, the corresponding Benefit Administration.

In relation to **AI quality control and auditing**. It will depend on:

• If it affects the exercise of the jurisdictional function, and, consequently, judicial independence, it corresponds, in accordance with the LOPJ, to the General Council of the Judiciary. Algorithmic surveillance of systems likely to affect judicial independence will require the CGPJ to collect and analyse data generated by algorithms to evaluate their performance, identify possible biases, errors or unwanted behaviour, and guarantee transparency and responsibility in the use of artificial intelligence.

• If it does not affect, it corresponds, in accordance with RD Law 6/2023, to the CTEAJE or the benefit administration, according to the scope of application of said system.

POLICY ON THE USE OF ARTIFICIAL INTELLIGENCE IN THE ADMINISTRATION OF JUSTICE

ANNEX I: DEFINITIONS

1. Artificial Intelligence (AI): System designed to operate with a certain level of autonomy and that, based on input data provided by machines or people, infers on how to achieve a set of established objectives using machine learning or logic and knowledge-based strategies, and generates information output, such as content (generative artificial intelligence systems), predictions, recommendations or decisions, that influence the environments with which it interacts.

2. Generative AI (GenAI): It refers to an artificial intelligence technology that derives new versions of text, audio or visual images from large amounts of data in response to user input. GenAI can be used in standalone applications, such as ChatGPT or Bard, or embedded in other applications. In addition, the administrations themselves can make systems of this type available. If you have any questions about what constitutes GenAI, please contact your administration's IT services.

3. "Hallucinations": It has become the term adopted by the GenAI community to describe how models will, from time to time, provide fictitious answers. The problem is not simply that the answers are wrong, but that they are safe and convincing. Society has developed an endemic automation bias, humans blindly favouring the suggestions of automated decision-making systems, often ignoring their own better judgment.

4. Cybersecurity: Cybersecurity is the practice of protecting computers, networks, software applications, critical systems, and data from potential digital threats. Organisations have a responsibility to protect data to maintain customer trust and comply with regulations.

5. Confidence and privacy: Confidential and sensitive information, including personal data of customers, employees or others, entered into publicly available GenAI applications, may leave residue within the model that may form part of a result elsewhere later, or be used to (re)train new models. Therefore, any personal information, proprietary or intellectual property information, or confidential information entered in the message may appear in the output of other users.

6. Model Bias: GenAI tools incorporate any bias from the data sets used to train them. This modelling bias does not always align with the core value of the Administration of Justice and its commitment to diversity, equity and inclusion. Therefore, the model output may make systematic errors or favour certain groups, leading to unfair or discriminatory results.

7. Intellectual Property: GenAI models are often trained on large publicly available data sets (for example, through data mining from public web pages). Therefore, the results may contain information protected by copyright or intellectual property of others. While ownership in many of these cases is unclear, users should err on the side of caution and not use any results that contain material that they suspect is under copyright protection in any material, internal or external.

8. Transparency Risk: All content generated using GenAI must be clearly identified in any externally facing content.

9. Third party risk: The data sent by administrations to third parties could be used in the use of GenAI tools by these actors external to the Administration of Justice. For example, sensitive user information is a potential risk if the organisation uses a third-party customer service chatbot provider that employs GenAI tools.

POLICY ON THE USE OF ARTIFICIAL INTELLIGENCE IN THE ADMINISTRATION OF JUSTICE

10. Algorithm: It is a set of defined instructions or rules that are used to perform a task or solve a specific problem. Algorithms are the basis of many AI systems, allowing machines to learn and make decisions.

11. Automatic Learning (Machine Learning): It is a subdiscipline of AI that focuses on the development of algorithms and models that allow computers to learn patterns and make predictions from data, without needing to be explicitly programmed for each task.

12. Artificial Neural Networks (ANN): They are computational models inspired by the functioning of the human brain that are used in machine learning and other fields of AI. They are made up of interconnected nodes (neurons) that process information and can adapt by adjusting the weights of the connections between them.

13. Deep Learning: It is a machine learning technique that uses artificial neural networks with multiple layers of processing to learn hierarchical representations of data. Deep learning has proven to be very effective in tasks such as image recognition, natural language processing, and autonomous driving.

14. Intelligent Robotics: It is a field of AI that combines artificial intelligence with robotics to develop robotic systems capable of perceiving their environment, making autonomous decisions, and performing tasks efficiently. Intelligent robotics are applied in a wide range of areas, from manufacturing to space exploration.

POLICY ON THE USE OF ARTIFICIAL INTELLIGENCE IN THE ADMINISTRATION OF JUSTICE

ANNEX II: PERMITTED AND PROHIBITED USE CASES

Any use of AI while performing work for employment-related matters with the Administration of Justice is subject to management approval. Next, examples of use cases that are subject to modification are shown. The following table does not reflect all possible use cases and scenarios. Use your best judgment and ask questions.

CATEGORY	USE CASES AND EXAMPLES	NECESARRY ACTIONS
Use cases generally permitted	Uses that do not involve personal or confidential information are only for an internal hearing and whose result will not be used in judicial or administrative resolutions. Examples: - Text translation from a secondary publicly available source. - Carry out an investigation of high-level background on a non-sensitive topic. - Generation of reports that do not affect rights or obligations of the citizen. - Get extensive document summaries, in the form of text, video or audio.	Always check if the service used is provided by the administration or another reliable organisation according to this protocol, to make a decision about its possible use. Review the results to make sure they fit the organisation guidelines. Remember that the data generated may not be accurate or complete, so results should be reviewed and reading the documents is not exempt if the norm so requires it. The data on which the AI works on cannot be used when its specifically personal or confidential data, but, also, remember that they may also contain information that may infer on the identification of people or in general, compromise information, through other data (Examples: the neighbour on X Street, 4ºB, Madrid. Or the neighbour of village X known as “Perico”).

POLICY ON THE USE OF ARTIFICIAL INTELLIGENCE IN THE ADMINISTRATION OF JUSTICE

Use cases that require the management's approval	Those who, in the previous case (without data with some type of protection), can be used to serve citizens, professionals or companies, with the data properly contributed by these people. Those who, in the previous case (without data with some type of protection), use information that will be used for coming to a decision. Examples: - Text generated for informative, promotional or formative material. - When the output is used in reports or documents, in general, whose recipients are citizens, professionals or companies. Those who, using private, personal or confidential data, or in general, data with some type of protection, can provide results that are necessary for the development and support of the activity that is specific to the organisation, being the service provided by the administration itself or specifically approved by the security department: Examples: - Certain reports that affect the privacy of people, or those who are treated in the scope of judicial investigations, such as textualisation in listening to telephone calls, their summaries, etc. - Support reports and help in jurisprudence issues, or similar. - Images and video generation using information. Could be used for reconstruction of scenarios in statements.	You must formally apply for the request to those responsible and obtain one or the explicit authorisations. You must inform the department of security, in addition to the administration in charge of innovation in AI, so that the service is published, if applicable, in the FAT records.
	Preparation of summaries and description of images and videos. It could be used to support analysis of difficult evidence perception by the human eye, such as tissue analysis, pattern detection in images, etc. - Generation of document drafts. - Simplification of legal texts to simple language.	
Cases that require the approval of IT	Generate a code to run on any device in the organisation.	Conduct a peer review of the code and contact the security department for its approval.

POLICY ON THE USE OF ARTIFICIAL INTELLIGENCE IN THE ADMINISTRATION OF JUSTICE

Prohibited	Use of personal data or with some type of protection whose use is not found in the previous cases. Specifically, use documents and data protected to generate information or documents for decision-making that affect the rights or obligations of citizens, professionals or companies, even though it is allowed with prior authorization, the generation of support documentation that is used in judicial records or in the decision-making sphere of the administration. Examples: - Obtain reports for making judicial or administrative decisions based on data provided by the user. - Generation of definitive resolutions that will be automatically issued.	Please report any prohibited use of the GenAI models to the security department
	Automated reporting of the status and details of the procedures to those interested in it, without previous human supervision. - Generation of documentation or support information about foreseeable future behaviour of people to base the judicial action based on their personal history or people of similar characteristics (e.g. not dictating a provisional detention or parole based on a report of AI recommendations)	
	The use of generative artificial intelligence systems that replace the irremediable presence on committees, meetings, or similar. Example: - Use of systems that process the audio of the meetings and perform automatic summaries can permit users to not have to attend physically.	

GOBIERNO DE ESPAÑA
MINISTERIO DE LA PRESIDENCIA, JUSTICIA Y RELACIONES CON LAS CORTES