Asian Journal of Social Science Studies; Vol. 10, No. 4; 2025
ISSN 2424-8517 E-ISSN 2424-9041
Published by July Press

AI and the Justice System: Approach From the EU AI Act

Susana Navas Navarro[1]

[1] Faculty of Law, Autonomous University of Barcelona, Barcelona, Spain

Correspondence: Prof. Susana Navas Navarro, Faculty of Law, Autonomous University of Barcelona, Barcelona, Spain.

Received: November 3, 2025 Accepted: December 5, 2025 Online Published: December 8, 2025

doi:10.20849/ajsss.v10i4.1539 URL: https://doi.org/10.20849/ajsss.v10i4.1539

Abstract

The use of artificial intelligence (AI) in the judicial system brings up big challenges, especially since basic legal rights are involved. In this paper, I first look at the rules in the European Artificial Intelligence Act, then discuss two main issues: First, does using AI change the way we think about legal mistakes, moving from just focusing on errors in the court system to a broader view of how the public service is working? And second, what kind of responsibility should the State have if someone gets hurt by AI used in the justice system? Finally, I explain the difference between high-risk and not high-risk AI-system in relation to fault-based and strict-based liability.

Keywords: artificial intelligence, judges, AI systems, State liability, public sector, risk-based, fault-based liability

1. Introduction

It is well known that the EU Commission has put forward a package of legislative proposals concerning civil liability for damages caused by AI systems. I am referring to the proposed Directives that was published on 28 September 2022 (COM[2022] 495 final and 496 final). One of these proposals resulted in the adoption of the Directive (EU) 2024/2853, of 23 October, on liability for damages caused by defective products allows Member States, which amended the still-in-force 1985 Directive (EUOJ L 18.11.2024. Hereinafter "Directive 2024"). The other remains a proposal, although work on it has resumed in an attempt to make progress on this much-needed regulation (Hacker, 2024).

The EU Commission has recently proposed to relax certain aspects of the AI Act (EUOJ L 2024/1689, 12.7.2024) through an omnibus proposal for regulatory simplification that affects several regulations governing the European digital market. This proposal delays the application of the rules relating to the requirements that high-risk AI systems implemented in the areas of Annex III of the AI Act must meet, as well as the obligations that the economic operators involved must comply with. Civil liability for damages caused by AI systems —beyond Directive 2024— will remain pending. At least until 2027, when the obligations of providers and those responsible for deployment enter into force, insofar as it appears to be the intention of the European legislator to take into account their non-compliance in order to determine a fault-based liability for damages caused by AI systems and models.

In this paper, I focus on civil liability arising from damages that may be caused by the implementation and development of AI systems in the public justice service. To do this, the first thing that must be explained is how these services fit within the AI Act. This Regulation sets out harmonized rules for placing AI systems on the market or putting them into service (Art. 1.2 AI Act), as well as for the use of AI systems within the European digital market; the prohibition of certain practices (Art. 5 AI Act); specific requirements for high-risk AI systems (Arts. 9–15 AI Act); and the obligations of operators of such systems—particularly the provider and the deployer (Arts. 16–27 AI Act). It also establishes certain rules on transparency (Art. 50 AI Act), on general-purpose AI models (Arts. 51–56 AI Act), and rules on market surveillance and enforcement of the AI Act itself (Arts. 72–84 AI Act). According to the AI Act, a "provider" is defined as "a natural or legal person, public authority, agency or other body that develops an AI system, or for whom an AI system is developed, and who places the AI system on the market or puts it into service under its own name or trademark, whether for payment or free of charge" (Art. 3.3). Article 3.4 defines the deployer as "a natural or legal person, public authority, agency or other body that uses an AI system under its own authority, except where the system is used in the course of a personal, non-professional activity".

The AI Act seeks to reduce administrative and financial burdens on companies, especially small and medium-sized enterprises (SMEs), including start-ups (Arts. 62–63 AI Act), taking into account a risk-based approach (Recitals Nos. 26–27 AI Act). AI tools used in the public justice service may be classified as high- or low-risk, and depending on that classification, certain requirements and obligations must be fulfilled. In addition, generative AI tools may also be used—for example, a general-purpose AI model such as a chatbot. However, in this paper I will focus on AI systems, not AI models, because it is AI systems that are being implemented in the public justice service and these are specifically addressed in Annex III of the AI Act.

The purpose of the AI Act is to ensure that when AI is used—including in the public justice service (which encompasses both the administration of justice and the delivery of justice in the strict sense)—certain safeguards or mechanisms are put in place to minimize risks to fundamental rights, safety, and the rule of law, among other possible risks (Art. 1.1 AI Act).

The AI Act aims to ensure trustworthiness when AI systems are used in the various areas it covers, including justice (Art. 6 and Annex III AI Act). AI systems can assist judges or other justice professionals in administering justice, as well as improve the efficiency of judicial processes. Nevertheless, national authorities retain the right to decide whether or not they wish to implement AI in the public justice service.

Due to the diversity of judicial systems in the EU and the varying level of AI development within them, I will take especially into consideration the application of AI systems in the public justice service in Spain, where the Royal Decree-Law 6/2023 of 19 December adopts urgent measures for the implementation, transformation, and resilience of the public justice service, the civil service, local government, and patronage (Spanish Official State Gazette No. 303, of 20 December 2023). This Decree remains in force after entering into force the Organic Law 1/2025 of 2 January, on efficiency measures in the public justice sector (Spanish Official State Gazette, No. 3, of 3 January 2025). The Royal Decree-Law 6/2023 regulates automated, proactive, and assisted judicial proceedings (Arts. 56–57). An “automated judicial proceeding” is “the judicial proceeding carried out by an information system that is properly programmed, without the need for human intervention in each individual case” (Art. 56.1). Proactive judicial proceedings are defined as “automated judicial proceedings self-initiated by information systems without human intervention, which make use of the information included in a file or procedure of a public authority for a specific purpose, in order to generate notices or direct effects for other, different purposes—within the same or other files—of the same or another public authority, in all cases in accordance with the law” (Art. 56.3).

Only the “assisted judicial proceedings” use AI systems, which may be high-risk. An “assisted judicial proceeding” is understood as “one for which the information system of the Administration of Justice generates a full or partial draft of a complex document based on data, which may be produced by algorithms and may serve as the basis for, or support of, a judicial or procedural decision” (Art. 57.1).

In Spain, the main robotization initiatives affect the transfer of information between different departments of the administration of justice, the cancellation of criminal records, the granting of nationality, the granting of pardons, and also the public prosecutor’s office and the Office of the State Attorney. AI is applied to document classification, document anonymization, automatic document processing, similarity analysis, easy-to-read text generation, and converting recordings and audio transcripts into text. Justice professionals have a search tool that integrates an AI system (*Delfos*) to support their work, as it helps with generating reports, searching and summarizing documents, and retrieving information, among other tasks. This has resulted in cost savings for the state budget of almost 16 million euros. On the other hand, AI4Justice is a virtual assistant that helps judges in Catalonia draft rulings (Gil, 2025).

In order to facilitate the implementation of AI tools in courts and prosecutors’ offices, the public administration prepared a document on the policy for the use of AI in the administration of justice (CTAJE, 2024).

As can be seen, the specific implementation of automation and AI in the public justice system corresponds to the three types of judicial proceedings referred to by the legislation mentioned above.

2. Provisions of the AI Act in Relation to the Justice System

Following the regulatory order of the AI Act itself, I will address, first, the prohibited practices (2.1.), then the high-risk AI systems (2.2.), next, those that are not high-risk (2.3.), and finally, the exclusions from classification as high-risk for certain AI systems (2.4.).

2.1 Prohibited Practices

The AI Act recognizes that certain practices involving the use of AI present excessive risks and therefore prohibits their use (Art. 5 AI Act). In the context of the public justice service, the AI Act prohibits placing on the market, putting into service, or using an AI system to carry out risk assessments of natural persons for the purpose of evaluating or predicting the likelihood that they will commit criminal offenses (“crime prediction,” “crime forecasting”), (Europol, 2024 and Yang, 2019) when such assessment is based exclusively on profiling that person or on evaluating their personality traits or characteristics (Art. 5.1 lit. d AI Act). The Commission was required to adopt principles for the practical implementation of this regulatory provision (Art. 96.1 lit. b AI Act). This practical guidance was adopted on 4 February 2025 (See online: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act. Date of Consultation: December 2025).

The basis for this prohibition lies in the presumption of innocence. Natural persons must always be judged on the basis of their actual behaviour. They must never be judged on the basis of behaviour predicted by an AI system that relies solely on profiling or on personality-related characteristics—such as nationality, place of birth, residence, number of children, level of indebtedness, or the type of vehicle they drive—without human assessment behind it and without a reasonable suspicion, based on verifiable objective facts, that the person is involved in criminal activity (Recital No. 42).

However, when it involves the application of the law to the extent that it is permitted by Union law or national law, the use of AI systems for these purposes would not be prohibited, although such systems would be considered high-risk systems, as I will explain below.

2.2 High-Risk AI Systems

The AI Act classifies certain AI systems used in specific areas—those described in Annex III—as high-risk (Art. 6.2). Placing them on the market and deploying them means that both providers and deployers are subject to strict compliance with a whole set of obligations (Arts. 16–27 AI Act) and that the AI systems meet a series of requirements (Arts. 8–15 AI Act). In addition, the obligations established in other European and national regulations must also be fulfilled—for example, personal data protection rules, defective product liability rules, or consumer protection rules, among others. Therefore, the AI Act does not exclude the application of those laws.

AI systems that judicial authorities intend to use to assist them in their functions—in searching for and interpreting facts and legal norms, as well as in applying the law to a specific set of facts—are considered high-risk systems (Annex III, section 8). This is due to their potentially significant impact on the rule of law and on the fundamental rights established in the EU Charter of Fundamental Rights, in particular the right to a fair trial, the presumption of innocence, effective judicial protection, judicial independence, human dignity (Gentile, 2022, Presno, 2022), and non-discrimination (Recital 7 AI Act). These rights are enshrined in the Charter of Fundamental Rights of the European Union (see online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12016P%2FTXT. Date of Consultation: December 2025) and are also referenced in the Council of Europe’s Ethical Charter on the Use of AI in Judicial Systems (European Commission for the efficiency of justice (CEPEJ), *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, Strasbourg, 3-4 December 2018. Online: https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment. Date of Consultation: December 2025). Any potential undermining of these rights has a direct impact on citizens’ trust in the justice system, the rule of law, and democracy itself.

The AI Act recognizes and reaffirms the role of the judge. AI may assist the judge and the public justice service in general, but it cannot replace the judge’s decision-making authority. The final decision must continue to be a human judicial decision—that is, a judge’s decision. Therefore, it makes clear that, for now, the so-called robot judge is not permitted. This point is also emphasized in the Council of Europe’s Ethical Charter on the Use of AI in Judicial Systems, which recognizes, for example, the citizen’s right to object and be heard by a human judge.

In Spain, Article 57 of Royal Decree-Law 6/2023 is especially clear in this regard with respect to so-called “assisted judicial proceedings,” in which the AI system assists the judge, magistrate, prosecutor, or court clerk but does not replace them (“…a complete or partial draft of a complex document based on data, which may be produced by algorithms…”). Specifically, Article 57(2), in its first sentence, warns: “Under no circumstances shall the draft document thus generated in itself constitute a judicial or procedural decision, without validation by the competent authority.”

That a robot judge might be possible in cases lacking complexity, as some authors suggest, is something I do not doubt (Cabrera, 2024, Sanch ś, 2023, Nieva, 2018, Laptev & Feyzrakhmanova, 2024, Susskind, 2019). However, since both the European and national legislators have excluded this possibility, I will not address it here.

On the other hand, as noted above, these systems are considered high-risk—insofar as they are permitted under national or EU law and meet the relevant requirements—when they concern the application of the law, according to Annex III, section 6. For the purposes relevant here, the following are of interest:

a) Systems used by authorities in the application of the law, or on their behalf by other bodies or institutions that support them, to assess the risk that a natural person will commit a criminal offense or reoffend, taking into account not only the profiling of natural persons or the assessment of personality traits and characteristics, but also past criminal behaviour of individuals or groups (Annex III, section 6 (d) AI Act).

b) AI systems intended to be used by law-enforcement authorities, or on their behalf by other bodies or institutions that support them, to create profiles of natural persons in the course of arrest, investigation, or prosecution of criminal offenses (Annex III, section 6 (e) AI Act).

In these cases—listed in a limited and precisely defined manner—their use is strictly necessary to achieve an essential public interest whose importance outweighs the risks. These systems may involve certain uses characterized by a significant imbalance of power and may result in surveillance, arrest, or deprivation of liberty of a natural person, as well as other negative effects on the fundamental rights enshrined in the European Charter. In particular, if the AI system is not trained with high-quality data, does not meet adequate requirements regarding performance, accuracy, or robustness, or is not properly designed and tested before being placed on the market or put into service, it may flag people in a discriminatory, incorrect, or unfair manner. Furthermore, it may hinder the exercise of important fundamental procedural rights, such as the right to effective judicial protection and an impartial judge, as well as the right to defence and the presumption of innocence—especially when such AI systems are not sufficiently transparent, explainable, or well-documented (Recital No. 59). This is why they are classified as high-risk systems.

2.3 Not-High Risk AI Systems

The classification of AI systems as high-risk should not be extended to AI systems intended for purely ancillary administrative activities that do not affect the administration of justice in specific cases, such as the anonymization or pseudonymization of judicial decisions, documents, or data, internal communication among staff members, or administrative tasks. These examples are cited in Recital 61 of the AI Act. To these, one may add assistance to citizens in their interactions with the public justice service (for example, through a virtual assistant). These fall under the category of “certain” AI systems (Art. 50 AI Act).

In this regard, among the judicial proceedings referred to in Articles 56 to 58 of the Royal Decree-Law, automated and proactive judicial proceedings would not be considered as produced by high-risk AI systems. It should be noted first that automation can be carried out without the involvement of AI. However, if AI were involved, they would still not be considered high-risk systems, even if they operate in a high-risk context established in Annex III of the AI Act. They would fall under the situation described in Article 6(3)(a) AI Act. Automation may occur primarily in civil, labour, and administrative-litigation proceedings, while in criminal proceedings it would be less evident, being limited to less serious offenses and expedited trials (Nieva, 2018; Barona, 2021, Pardo 2024).

2.4 Exclusions

The AI Act establishes that certain AI systems applied in the judicial context—systems that, *prima facie*, could be classified as high-risk—are nevertheless excluded from that classification because they do not present a significant risk of causing harm to the fundamental rights of those involved in judicial proceedings, as they do not exert a material influence on the outcome of the judicial decision. The use of these systems, excluded from the high-risk category, may still be relevant in the field of justice.

The exclusion from the classification applies in any of the following cases:

a) When the AI system is intended to carry out a limited procedural task, such as transforming unstructured data into structured data, categorizing documents, or detecting duplicates. The risks they pose are limited, even if the system is used in a context listed in Annex III as high-risk, such as the public justice service.

b) When the AI system is intended to improve the output of a previously performed human activity—for example, AI systems designed to improve the language of a document drafted by a human.

c) When the AI system is intended to detect decision-making patterns or deviations from previous decision-making patterns and is not intended to replace the previously performed human assessment or influence it. For example, systems that detect possible inconsistencies or anomalies.

d) When the AI system is intended to carry out a preparatory task for an assessment relevant to the use cases listed in Annex III, such that the impact of its output is minimal in terms of posing a risk to the subsequent assessment. This includes, for example, intelligent solutions for file management, indexing, searching, text and speech processing, linking data to other data sources, or document translation.

Nevertheless, AI systems referred to in Annex III are always considered high-risk when the AI system performs profiling of natural persons.

To guide the application and interpretation of the AI Act, the Commission is empowered to adopt delegated acts establishing a list of uses of high-risk systems and uses of systems that are not high-risk, as well as to monitor whether any system excluded from this category should be added to it (Art. 7 AI Act). These guidelines must be published no later than February 2026 (Art. 6(5) AI Act). This guidance is particularly relevant for those Member States that plan to implement AI in the justice system or are already using it. Based on available information, projects to implement AI use in national justice services are being launched. Six Member States (Germany, Austria, Portugal, Spain, Luxembourg, and France) have already been using AI applications both in courts and in public prosecutors' offices for activities relevant to their functions since 2023 (Communication from the Commission to the European Parliament, the Council, the European Central Banck, the European Economic and Social Committee and the Committee and the Committee of the Regions. 2024 EU Justice Scoreboard, COM/2024/950 final.
https://public.tableau.com/app/profile/cepej/viz/ResourceCentreCyberjusticeandAI/AITOOLSINITIATIVESREPORT?publish=yes. Date of Consultation: December 2025).

Depending on the type of AI system involved, the public authority, whether acting as provider or deployer, must comply with a range of obligations. Non-compliance may result not only in regulatory sanctions as outlined in the AI Act but also in potential state liability if harm is caused.

With respect to the use of high-risk AI systems in the public justice service, two key issues arise, which will be explored in the following two sections.

3. Judicial Error or Abnormal Functioning of the Public Justice Service? The Case of Spain

The first issue concerns whether the implementation of AI-based tools to support judicial authorities — that is, high-risk systems — can, when malfunctioning and causing harm to third parties, be classified as a judicial error or as an instance of abnormal functioning of the public justice service. It appears clear that low-risk systems applied to purely administrative tasks, where fundamental rights are not at stake, fall within the domain of abnormal service functioning if — and only if — harm is indeed caused to third parties.

Nevertheless, the issue becomes less straightforward when dealing with high-risk systems that assist judicial authorities in interpreting facts, applying the law to a given set of facts, or proposing draft decisions, or in other judicial activities involving fundamental rights of citizens, such as the right to effective judicial protection. In such cases, the fact that AI systems assist judges — who must validate, review, or reject the output generated — implies that, if harm results from, for example, an erroneous interpretation of facts or legal norms produced by the AI system, it might be attributable to judges, since the final decision ultimately rests with them. In such circumstances, the situation could be deemed a judicial error (Gutiérrez, 2023).

In Spain, Article 292 of the Organic Law 6/1985, of 1 June, of the Judiciary (Spanish Official State Gazette, No. 157, 2 July 1985) differentiates judicial error from the abnormal functioning of the administration of justice with respect to the requirements established for one or the other to occur and to activate the civil liability of the administration. In the first case, a prior judicial declaration of the judge's error is necessary before initiating a compensation claim process by the victim. In contrast, in the second case, a prior judicial declaration of malfunctioning of the administration of justice is not required. From the perspective of the victim who has suffered damages due to a wrongful judicial decision, the path of judicial error leaves them somewhat unprotected compared to if the same facts were considered a malfunction of the administration of justice.

Indeed, the concept of *judicial error*, as an "indeterminate legal concept" (Gutiérrez, 2023), is generally understood to refer to the incorrect assessment of facts or misapplication of the legal system made in a judicial decision during the entirely human exercise of judicial activity.

When an AI system is involved in judicial activity, and its implementation or deployment has been decided by the justice authority to organize its internal service with the aim of increasing efficiency, responsibility for the AI system's outcomes that may cause harm does not rest solely on the judge's evaluation. It also depends on the public authority's compliance — whether as provider or as deployer (Articles 8 et seq AI Act) — with a series of obligations, including, for example, human oversight to detect potential system malfunctions.

Furthermore, the judge may be unaware of the AI system's internal workings (i.e., the specific instructions programmed), the data on which it was trained, or whether the administration, as provider or deployer, has fulfilled its obligations. Additionally, the opacity of algorithms can hinder understanding of how results are generated (Burrell, 2016, Palmiotto, 2012).

This is so despite efforts to ensure that AI systems, or AI-based tools in general, are transparent (Article 13 AI Act), a basic requirement to maintain the integrity of the public justice service.

AI systems are typically trained on historical data that may reflect discriminatory biases unknown to the judge (Hacker, 2018). As mentioned earlier, judges are only provided with the necessary operational instructions by the public authority for handling the AI system. There is a distinction between explainability — why a particular algorithmic decision was made — and its comprehension by the personnel involved (Cantero, 2024). The latter requires that the explanation be correctly interpreted to enable an informed decision, as a judge must do when using AI tools that assist in judicial decision-making by presenting draft rulings.

In practice, however, as previously noted, judges may be limited to receiving instructions without truly understanding the internal functioning of AI systems. Even if the system is transparent and its decisions are explained, comprehension is a different matter altogether.

Appealing to explainability and transparency can create a false impression that the judge is responsible for having committed an "error" when actually despite the system being public, the technical understanding of the system may be difficult. Therefore, I contend that the implementation of AI within the public justice service should shift, in Spain, responsibility for judicial error from judges to the abnormal functioning of the justice service. This concept is understood as: "malfunctions in the operation of courts and tribunals as an organic whole, integrating various individuals, services, resources, and activities" (Gutiérrez, 2023). AI, viewed as a tool or service, could fit within this framework.

In any case, where harm may result from a judgment with which the citizen disagrees, and whose draft was proposed by an AI system, if the individual believes that the involvement of the AI system caused him to lose a chance, it should be borne in mind that, as long as the judicial decision is not final, the appropriate and primary course of action should be an appeal before the superior court. Only when no further appeal is possible could the exercise of a claim for liability, in this case for loss of chance, be admissible.

Therefore, such a claim must be subsidiary (Medina, 2007). The inherently contingent nature of this doctrine means that, upon appeal, a different court may issue a decision that renders the liability claim unnecessary, insofar as the claimant no longer considers that they lost the chance.

Moreover, as the claim for damages liability is subsidiary until the case reaches final instance, the court's infringement must be, in accordance with the Court of Justice of the European Union's (CJEU) decision in the Köbler case (C-224/01, Gerhard Köbler v. Republic of Austria [2003] ECR I-10239), "manifestly serious". Ultimately, this aims to prevent the proliferation of unfounded lawsuits against judges and courts and to ensure legal certainty. In fact, there are Member States in which citizens are *de facto* or *de jure* prohibited from bringing claims based on judicial errors (Scherr, 2011). Hence, the importance of introducing a rule regarding the evidence, which I will refer to later, so that a potential claimant can verify whether there is a serious basis for pursuing their claim.

4. Fault- or Strict-Based Liability in Case of Damages Due to the Implementation of AI in the Justice System?

The second issue concerns which liability regime is most appropriate when harm results from the implementation of AI in the public justice service: a fault-based or a strict liability regime?

Regarding the liability of public authorities, not all EU Member States provide for a strict liability regime; many apply a fault-based liability standard (Oliphant, 2016). In the case of Spanish law, for example, although the legislation appears to establish strict liability (Law 40/2015 of 1 October on the Legal Regime of the Public Sector [Spanish Official State Gazette, No. 236, 2. October 2015]), in practice the courts have required proof of the negligent conduct by the public agent (Mendilibar, 2024) — negligence that must be demonstrated by the injured party. Therefore, despite the legal façade of a strict liability regime, a fault-based regime is actually applied (Mendilibar, 2024).

Indeed, this is not the only fault-based liability system masquerading as a risk-based one. For instance, the liability regime applicable to defective products is often presented as strict liability, but in reality, it relies on fault. Not because negligence must be proven in the traditional sense, but because the claimant must demonstrate that certain duties incumbent upon the manufacturer—such as duties of care or information—and compliance with technical standards have been breached. Such regimes have sometimes been described as "quasi-strict" liability, while in fact they are fault-based under a different guise (Zech, 2021, Wagner, 2017). Only in cases of manufacturing defects is it conceivable that a truly strict liability rule exists.

Starting from a fault-based liability regime, it is evident that when a high-risk AI system, as defined in the paragraph 8 of Annex III of the AI Act, causes harm to a citizen, the general rules regarding the burden of proof do not confer any special advantage on the injured party.

At this point, it is necessary to reference the Report of the Expert Group on Liability and New Technologies (NTF) (*Liability for Artificial Intelligence and other emerging technologies*, 29 November 2019, online: https://op.europa.eu/en/publication-detail/-/publication/1c5e30be-1197-11ea-8c1f-01aa75ed71a1/language-en. Date of consultation: November 2025). This Report starts from the general rule that the victim must prove the cause of the harm but acknowledges that the complexity of the technology involved may give rise to an information asymmetry between the responsible operator and the victim, making it impossible or excessively burdensome for the latter to prove causation. Therefore, in Recommendation no. 26, it lists a series of circumstances that would justify the European legislator, or even national legislators, adopting a general rule for the reversal of the burden of proof concerning the causal link.

These circumstances include: the likelihood that the technology contributed to causing the harm; the probability that the harm was caused either by the intervention of the technology or by another factor within the same sphere of control; the risk of a known defect in the technology even if its impact on causation is not evident; the degree of ex post traceability and intelligibility of the AI-governed processes that may have contributed to the harm (information asymmetry); the extent of subsequent access to and understanding of the data collected and generated by the technology; and the type and extent of potential and actual harm caused.

Additionally, in Recommendation no. 24, the Report suggests presuming causality whenever there is a detected breach of safety standards whose observance would have prevented the harm. According to the Expert Group on Liability and New Technologies (NTF, 2019), these mechanisms would complement the rebuttable presumption, according to which causality exists unless it is possible to identify the actors who manipulated the device.

Given the asymmetry that exists between the victim and the public authority—whether as provider or as a deployer of the system—and the potentially extraordinary difficulties in proving certain elements of liability, such as causation, system malfunction, or breach of the obligations incumbent upon the public authority under the AI Act, the Proposal for a Directive on Non-Contractual Liability for Damage Caused by AI Systems dated 28 September 2022, while not a panacea and containing debatable aspects and areas for improvement (Hacker, 2024, Brune et al., 2022), better balances the interests of the parties in line with the aforementioned Report by presuming *iuris tantum* the causal link between the defendant’s fault and the outcome produced by the system in the case of high-risk AI systems (Article 4). Indeed, the Proposal establishes several presumptions aimed at alleviating the burden of proof on the victim.

Furthermore, the regulation of a right of access to relevant information (Article 3 of the Proposal for a Directive) may contribute to improving the position of the claimant or potential claimant regarding evidentiary matters, taking into account that the information should be presented in a manner that is easily comprehensible to the victim.

This could therefore represent a significant advance. However, one may question whether, given the fundamental rights at stake—as previously indicated—it would not be advisable to introduce a genuine regime of strict liability for those actions involving the use of high-risk AI systems. In the case of Estonia, for instance, it has been asserted that the distinction between these actions should be irrelevant when applying the liability regime, which in any case should be risk-based and not fault-based (Pilving, 2023).

An important element to consider is the case law of the Court of Justice of the European Union (CJEU) concerning state liability, which has made clear since the landmark Francovich judgment (Cases C-6/90 y C-9/90. Andrea Francovich and Danila Bonifaci and others v. Italian Republic [1991] ECR I-5357) —where Italian workers sued the Italian Republic—that state liability is based on risk and adds that national rules on compensation cannot be less favourable (Craig, 1997). In the Brasserie du Pêcheur/Factortame cases (C-46/93, C-48/93), the Court held that fault was not a relevant factor in assessing the breach of EU law.

It is true that this case law focuses on the State’s failure to comply with European legislation that granted rights to its citizens, which were subsequently denied. In other words, it is not precisely the same situation as the one under consideration here. However, a more analogous, though not identical, case is the Köbler judgment, which deals with a breach of EU law by a national court, thereby constituting an erroneous judicial decision. This case highlights the three elements for such an erroneous judicial decision to give rise to liability: (1) the breached rule;

(2) the breach must be manifestly serious; and (3) the causal link between the breach of the obligation incumbent on the Member State and the damage suffered by the victim. At no point is fault mentioned as a basis for liability. Rather, the CJEU avoids referring to it (Scherr, 2011). However, the standard of seriousness is heightened compared to the Francovich case, where a "sufficiently serious" breach was enough.

The requirement of a "manifestly serious" breach must not be equated with the need to prove fault, as emphasized by the CJEU in the case Traghetti del Mediterraneo SpA v. Italy (C-173/03). This jurisprudence has even led to assertions that state liability rules based on fault should be amended to adopt a regime of liability based on risk (Scherer, 2002).

Case law to the same effect—that is, recognizing a regime of strict liability—has been established particularly in relation to discrimination cases (Marshall, C-271/91, Arjona Camacho C-407/14, DX C-113/22) (Navas, 2008), a notably sensitive issue when it involves high-risk systems that are fed with large volumes of data, within which biases may exist and subsequently be reproduced in the system's output.

Therefore, drawing upon or inspired by this jurisprudential doctrine, it can be affirmed that the CJEU advocates for a risk-based liability regime when the party causing the damage is the State or a public agent acting on its behalf.

Based on the foregoing, revisiting the Proposal for a Regulation accompanying the European Parliament Resolution of 20 October 2020 (Resolution 2020/2014[INL] of the European Parliament of 20 October 2020; Report COM/2020/64 final of the Commission of 19 February 2020, https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1593079180383&uri=CELEX%3 A52020DC0064. Date of Consultation: December 2025) may be an option. This would mean that the regulation would be directly applicable, thereby avoiding the greater or lesser degree of disharmony that could arise from opting for a Directive (Frizberg, 2024). Such a legislative policy decision would break with the EU's tradition of respecting the distinct tort liability systems of each EU Member State in matters of non-contractual liability.

Nonetheless, the Commission has not hesitated to transform Directives into Regulations, given the fragmented outcomes that the implementation of the former has generated as exemplified by the Directive on data protection, the General Product Safety Directive [Regulation (EU) 2023/988 of the European Parliament and of the Council of 10 May 2023 on general product safety is applicable (EUOJ L 135, 23 June 2023)], and the so-called "Machinery Directive. [Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery (EUOJ L 165, 29 June 2023)]. All of these are now Regulations. Furthermore, the regulation of the European digital market is increasingly carried out through Regulations, particularly the Digital Services Act [Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (EUOJ L 277, 27 October 2022)] the Digital Markets Act [Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector (EUOJ L 265, 12 October 2022)] and, of course, the AI Act and the Data Governance Act. Crucial fundamental rights are at stake.

The question that arises from these considerations is whether the future Liability Directive Proposal—whether it remains a directive or eventually becomes a regulation—could allow for a pure strict liability regime for high-risk AI systems. Opting for such a regime would simplify the claims process for damages, as victims would no longer need to prove a breach of duty or a malfunction of the AI system. It would reduce costs and complexity (Wachter, 2024; Karner, Koch, Geistfeld, 2021, Rub í, 2024, Navarro-Michel, 2020). Moreover, it would ensure that those who benefit from an activity that generates risks are the ones who bear the costs of any resulting harm.

However, a pure strict liability regime also presents drawbacks. It could significantly hinder technological innovation, especially for SMEs and start-ups. Such a barrier could lead to a loss of investment in AI-based products that may bring substantial benefits to society and individuals, such as new medical devices, predictive medicine, or autonomous vehicles capable of drastically reducing accident-related harm and saving lives (Hacker, 2024, Mayrhofer, 2024). Additionally, if AI system providers must assume the full cost of all potential damages their systems may cause, this could increase the price of such systems, ultimately restricting access to them to only certain companies or institutions and excluding part of the population from their benefits. Finally, in cases involving non-material (moral) damages, a strict liability regime may encourage opportunistic claims, given the inherent difficulties in proving such harm, which is often harder to observe and verify than material damage (Brune et al., 2022).

In light of this, it is worth asking whether a fault-based system, such as that in the Liability Directive Proposal, or a quasi-strict liability system, like that of the 2024 Directive, might be more effective by offering a better balance between fostering AI innovation and ensuring compensation for the harm it may cause. In this regard, due to the asymmetry that exists between the provider of AI systems and models and the victim, both legal texts include—as previously mentioned—mechanisms that allow access to relevant information. These mechanisms enable the victim, whether a potential claimant or an actual claimant, to make an informed decision about whether to file a claim. This access to relevant information may also have the indirect effect of encouraging providers to comply with their obligations under the AI Act (Hacker, 2024).

Nevertheless, I consider it important to recall that not all automated judicial proceedings within the public justice service carry the same level of risk, as I have previously discussed. In this regard, in my opinion, this distinction should be taken into account when determining the applicable liability rules (Pilving, 2023). Therefore, when the systems in question are not high-risk, the rules of a fault-based liability could apply, whereas for high-risk systems, an objective (strict) liability regulation seems more than appropriate in order to protect citizens against violations of fundamental pillars of the democratic state governed by the rule of law, as previously emphasized. This is, ultimately, the standards foreseen in the Regulation accompanying the European Parliament Resolution of 20 October 2020. Moreover, in the case of low-risk systems, Article 8 of this Regulation reverses the burden of proof in favour of the victim, whereas the provision in the Proposal for a Directive on Non-Contractual Liability for Damages Caused by AI systems leaves it to the judge’s discretion whether to apply the presumption of causation or not, which results in greater legal uncertainty (Brune et al., 2022).

In any case, the rule on the production of evidence in Article 3 of the Proposed Directive should be contemplated under both types of liability —both the risk-based for damages caused by high-risk AI systems and the fault-based for those caused by low-risk systems.

In contrast, the 2020 Proposal for a Regulation did not provide for any rebuttable presumptions or rules on the reversal of the burden of proof, whether concerning the causal link or the malfunctioning of the AI system. It left to the discretion of the Member States, through the adjustments they might make to their respective legal systems, whether to adopt all, some, or none of such presumptions or rules.

Starting from the European legislator’s preference for rebuttable presumptions as a legal means to address causal uncertainty, the Proposal for a Directive on Non-Contractual Liability for Damages Caused by AI systems or the rules established in the Directive (EU) 2024/2853, of 23 October, on liability for damages caused by defective products allows Member States (Article 10) can serve as an inspirational reference. These set forth rules intended to alleviate the victim’s burden of proof, and as mentioned, the legislator could draw on these provisions—appropriately adapted—to incorporate them into a strict liability regime in cases of damages caused by high-risk AI systems. For instance, presumptions of causation could arise where the defendant has failed to produce evidence they were obligated to present; where the claimant demonstrates that the AI system fails to comply with mandatory requirements under the AI Act. These are examples of circumstances that could justify such presumptions.

On the other hand, a mere statement by the provider that the system or AI model may produce defective or fallible results cannot, by itself, serve as a means of exonerating liability.

Finally, the exclusion of the application of the public authority’s exoneration based on the development risks should be considered in these cases. It must be borne in mind that the implementation of AI is a public authority decision concerning the internal organization of its services aimed at achieving greater efficiency. These services could be performed by a human being perfectly well. In fact, this has been the practice until recently. Therefore, the burden of bearing the damage cannot be placed on the citizen by invoking the state-of-the-art at the time the harmful events occur, since those events could have arisen from human behaviour, in which case the exception of the state-of-the-art could not be invoked.

In any event, in the context of the public justice service, given that a “robot judge” is not admissible and the results generated by the AI system must be supervised by a human, it does not appear that the requirements for applying the exoneration of liability by the public authority are met, particularly in the judicial sphere.

Furthermore, it should be noted that Directive (EU) 2024/2853, of 23 October, on liability for damages caused by defective products allows Member States, when incorporating it into their national laws, to exclude the exoneration based on development risks in accordance with certain parameters established in Article 18.

To conclude, I would add that the liability regime should be unified, that is, applicable equally to both public and private actors. In fact, the recital of the Directive Proposal explicitly warns that it applies to State liability since public authorities are also covered by the provisions of the AI Act as they are subject to the obligations set therein.

5. Conclusions

The main conclusions of this work are as follows. First, the AI Act treats high-risk AI systems used in the justice sector as tools that assist judges, not as replacements. Second, not every AI system used in the public justice system is considered high-risk—some are used in ways that do not threaten fundamental procedural rights. Third, when a high-risk AI system helps a judge with tasks like examining evidence, interpreting facts, or applying the law, any harm that results should be seen as a problem with how the public justice service is functioning, not as a traditional judicial error. Fourth, when high-risk AI systems cause harm, the public justice service should be held liable based on risk, while for low-risk systems, liability should depend on fault. This framework should be supported by the disclosure rules proposed in the 2022 Directive, along with rebuttable presumptions that make it easier for victims to prove causation and that the AI system malfunctioned.

In the end, suing for damages caused by a wrong judicial decision should only be an option after all appeals have been completed up to the highest court. And if the Court of Justice of the European Union's Köbler doctrine is applied, a judge or court must have committed a "manifestly serious" violation—without needing to prove fault. This prevents a flood of baseless claims against judges. For this reason, it would be wise to adopt a rule requiring disclosure of evidence, so that a claimant can check whether their case actually has a solid and serious foundation before filing a lawsuit.

References

Barona Villar, S. (2021). *Algoritmización del Derecho y de la justicia*. Valencia: Tirant Lo Blanch.

Brune, J., *et al.* (2022). The European Commission's Approach To Extra-Contractual Liability and AI – A First Analysis and Evaluation of the Two Proposals. *CiTiP Working Paper.* KU Leuven Centre for IT & IP Law (1-64).

Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 1-12. https://doi.org/10.1177/2053951715622512

Cabrera Fernández, M. (2024). Aplicación de la inteligencia artificial a la toma de decisiones judiciales. *Eunomía. Revista en Cultura de la Legalidad*, *27*, 183-200. https://doi.org/10.20318/eunomia.2024.9006

Cantero Gamito, M. (2024). Acceso a la justicia en tiempos de IA: ¿hacia una justicia low-cost?. *Revista CIDOB d'afers internacionals*, *138*, 65-78. https://doi.org/10.24241/rcai.2024.138.3.51

Comité Técnico Estatal de la Administración de Justicia Electrónica (CTEAJE). (2024). Política de uso de la inteligencia artificial en la administración de justicia. Secretaría general del CTAJE.

Council of Europe, CEPEJ. (2018). *European ethical charter on the use of artificial intelligence in judicial systems and their environment*. Strasbourg. Retrieved December 2025, from https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment

Craig, P. P. (1997). Once more upon the Breach: The Community, the State and Damages Liability. *LQR*, *113*, 67-78.

European Commission. (2024). *EU Justice Scoreboard 2024* (COM/2024/950 final). Retrieved December 2025, from https://public.tableau.com/app/profile/cepej/viz/ResourceCentreCyberjusticeandAI/AITOOLSINITIATIVESREPORT?publish=yes

European Parliament and Council. (2024). Directive (EU) 2024/2853 on liability for defective products. *Official Journal of the European Union.*

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. *Official Journal of the European Union.* Retrieved December 2025, from http://data.europa.eu/eli/reg/2024/1689/oj

Europol. (2024). A*I and policing the benefits and challenges of artificial intelligence for law enforcement*. An Observatory Report from the Europol Innovation Lab, 23 September.

Fenoll Nieva, J. (2019). *Inteligencia artificial y proceso judicial*. Madrid: Marcial Pons. https://doi.org/10.2307/jj.26844203

Frizberg, D. (2024). Adapting liability rules to artificial intelligence. EPRS. European Parliament.

Gentile, C. (2022). AI in the courtrooms and judicial independence: An EU perspective. *EUIdeas*. https://doi.org/10.2139/ssrn.4198145

Gil Seaton, A. (2025). Eficiencia organizativa y uso de la IA en los tribunals españoles. Judicial Efficiency revisited. Retrieved December 2025, from https://iaplcolloquium2025.pravo.hr/general-information

Gutiérrez Santiago, P. (2023). La responsabilidad por daños derivados de errores judiciales en materia “civil”: de su aparente laxitud legal a su extraordinariamente restrictiva concepción jurisprudencial en España. In García Amado, J. A. *et al.* (Eds.), *El error judicial. Problemas y regulaciones* (pp. 152-183). Valencia: Tirant Lo Blanch.

Hacker, P. (2018). Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law. *Common Market Law Review, 55,* 1143-1186. https://doi.org/10.54648/COLA2018095

Hacker, P. (2024). *Proposal for a Directive on adapting non-contractual civil liability rules to artificial intelligence: A complementary impact assessment*. European Parliament. Retrieved December 2025, from https://www.europarl.europa.eu/RegData/etudes/STUD/2024/762861/EPRS_STU(2024)762861_EN.pdf

Karner, E., Koch, B. A., & Geistfeld, M. A. (2021). *Comparative Law Study on Civil Liability for Artificial Intelligence*. Comisión europea. https://doi.org/10.1515/9783110775402

Laptev, V. A., & Feyzrakhmanova, D. R. (2024). Application of artificial intelligence in justice. *Human-Centric Intelligent Systems*, *4*, 394-40. https://doi.org/10.1007/s44230-024-00074-2

Mayrhofer, A.-C. (2024). Product Liaiblity in the age of AI – Proposal for a “two track” solution. *Revista electrónica de Direito*, *33,* 106-130. https://doi.org/10.24840/2182-9845_2024-0001_0005

Medina Alcoz, L. (2007). *La teoría de la pérdida de oportunidad. Estudio doctrinal y jurisprudencial de derecho de daños público y privado*. Cizur Menor: Thomson-Civitas.

Mendilibar Navarro, P. (2024). *Determinación de la responsabilidad patrimonial de la administración en la toma de decisiones basadas en inteligencia artificial*. Valencia: Tirant Lo Blanch.

Navarro-Michel, M. (2020). Vehículos automatizados y responsabilidad por producto. *Revista de Derecho civil, II*(5), 215-216.

Navas Navarro, S. (2008). El principio de no discriminación en el derecho contractual europeo. *ADC. LXI*(III), 1475-1485.

Oliphant, K. (2016). *The liability of public authorities in comparative perspective, European Group on Tort Law.* Cambridge University Press. Retrieved December 2025, from https://www.cambridge.org/core/books/liability-of-public-authorities-in-comparative-perspective/liability-of-public-authorities-in-comparative-perspective/154A779CB5CA8B6A04AD1DF48FFADCD9

Palmiotto, F. (2021). The black box on trial: The impact of algorithmic opacity on fair trial rights in criminal proceedings. In M. Ebers, & M. Cantero Gamito (Eds.), *Algorithmic governance and governance of algorithms: Legal and ethical challenges* (pp. 49-70). Springer. https://doi.org/10.1007/978-3-030-50559-2_3

Pardo Iranzo, V. (2024). Los principios del procedimiento en tiempos de justicia digital e inteligencia artificial: la regla de la automatización y sus consecuencias. In S. Calaza López, & I. Odeñana Gezuraga (Eds.), *Next Generation Justice: Digitalización e Inteligencia Artificial* (pp. 130-131). La Ley.

Pilving, I. (2023, February 20). Guidance-Baed Algorithms for Automated Decision-Making in Public Administration: The Estonian Perspective. *CERIDAP*. Retrieved December 2025, from https://ceridap.eu/guidance-based-algorithms-for-automated-decision-making-in-public-administration-the-estonian-perspective/?lng=en

Presno Linera, M. A. (2022). *Derechos fundamentales e inteligencia artificial*. Madrid: Marcial Pons. https://doi.org/10.2307/jj.4908196

Retrieved December 2025, from https://www.europarl.europa.eu/RegData/etudes/STUD/2024/762861/EPRS_STU(2024)762861_EN.pdf

Rubí Puig, A. (2024). Inteligencia artificial y daños indemnizables. *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial*. APDC. Aranzadi: Cizur Menor (pp. 621-688).

Sanchís Crespo, C. (2023). Inteligencia artificial y decisiones judiciales: crónica de una transformación anunciada. *Scire*, *29*(2), 65-80. https://doi.org/10.54886/scire.v29i2.4937

Scherer, A. (2002). “State Liaiblity – Ten years after Francovich”. Retrieved December 2025, from https://www.lunduniversity.lu.se/lup/publication/1554589

Scherr, K. M. (2011). Comparative aspects of the application of the principle of State liability for judicial breaches. *ERA Forum*, *12*, 565 ff. https://doi.org/10.1007/s12027-011-0242-8

Susskind, R. (2019). *Online courts and the future of justice*. Oxford University Press. https://doi.org/10.1093/oso/9780198838364.001.0001

Wachter, S. (2024). Limitations and loopholes in the EU AI Act and AI Liability Directives: what this means for the European Union, the United States, and beyond. *Yale Journal of Law and Technology, 26*(3), 671-717. https://doi.org/10.2139/ssrn.4924553

Wagner, G. (2017). Produkthaftung für autonome Systeme. *AfC*, *217*(6), 712. https://doi.org/10.1628/000389917X15126388934364

Yang, F. (2019). Predictive policing. In *Oxford Research Encyclopedia: Criminology and Criminal Justice*. Oxford University Press.

Zech, H. (2021). Liability for AI: public policy considerations. *ERA Forum*, *22*, 150. https://doi.org/10.1007/s12027-020-00648-0

Copyrights