

ADELE

Ethical self-assessment

Table of Contents

1. Scope of the assessment	3
2. Relevant ethical standards	3
1) High-Level Expert Group’s Ethics Guidelines for Trustworthy Artificial Intelligence	5
2) European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment	6
3. Compliance measures	8
1) Human agency and user control	8
2) Respect of fundamental rights in judicial proceedings	9
3) Fairness and non-discrimination	10
4) Transparency	11
5) Technical safety	12
6) Accountability	12
4. Conclusions.....	13

1. Scope of the assessment

This document provides an ethical self-assessment of the research activities conducted in the ADELE project. In particular, this assessment concerns the activities pertaining to the development of the AI-powered pilot tool¹. Legal issues pertaining to personal data and GDPR compliance have already been detailed in D2.3 (Final annotated corpus).

2. Relevant ethical standards

The utilization of AI applications has experienced tremendous growth in recent years, bringing forth numerous benefits and conveniences. However, this expansion has also provoked ethical concerns, such as privacy breaches, algorithmic discrimination, security and reliability issues, transparency, and other unintended consequences. There are over 200 governance policies and ethical guidelines for AI usage published by different public bodies, academic institutions, professional bodies, standards-setting bodies, private companies, and civil society organizations around the world. No matter their form, be they ethics codes, guidelines, frameworks, or policy strategies, they all embed and evolve around highly abstract principles that provide little protection from potential harms related to AI as they rarely offer any guidance on how to design and deploy algorithms within these ethical boundaries.

During the project lifetime, we have reviewed a significant number of these documents and found several to be containing more explicit sets of ethical standards relevant to the development of AI applications for the judiciary:

1. **Asilomar AI Principles** released by the Future of Life Institute² (2017);
2. **European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment** adopted by The Council of Europe acting through the European Commission for the Efficiency of Justice (CEPEJ)³ (2018);
3. **White paper AI Governance: A holistic approach to implement ethics into AI** released by the World Economic Forum (WEF)⁴ (2019);

¹ <https://adele-tool.eu/>.

² Future of Life Institute, Asilomar Principles, 2017, <https://futureoflife.org/open-letter/ai-principles/>.

³ Council of Europe, European Commission for the Efficiency of Justice (CEPEJ). European ethical charter on the use of artificial intelligence in judicial systems and their environment. 2018. available online at <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>.

⁴ World Economic Forum, AI Governance A Holistic Approach to Implement Ethics into AI, White Paper, 2019, <https://www.sipotra.it/wp-content/uploads/2019/05/AI-Governance.-A-Holistic-Approach-to-Implement-Ethics-into-AI.pdf>.

4. **AI4People’s Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations** by the Atomium – EISMD (AI4Poeple)⁵ (2018);
5. **EU Ethics Guidelines for Trustworthy AI by the European Commission through the High-Level Expert Group on Trustworthy AI (HLEG)**⁶ (2019);
6. **Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector** issued by the Alan Turing Institute⁷ (2019);
7. **Preliminary Study on the Ethics of Artificial Intelligence** issued by UNESCO (2019) and first draft of the Recommendation on the Ethics of Artificial Intelligence by the UNESCO Ad Hoc expert Group (AHEG)⁸ (2020);
8. **Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment** released by the Independent High-Level Expert Group on AI⁹ (2020);
9. **Guidance Note on Ethics by Design and Ethics of Use Approaches for Artificial Intelligence** by a panel of experts at the request of the European Commission DG Research & Innovation¹⁰ (2021);
10. **OECD Recommendation of the Council on Artificial Intelligence**¹¹ (2022);
11. **Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)**¹² (2021, as further amended in 2023 by the European Parliament).

The former represents the first institutional document issued by the EC containing relevant ethical standards for trustworthy AI. This document has inspired subsequent discussion on AI in the EU and around the world and has contributed to a regulatory debate on AI that has led to the adoption of the AI Act proposal. Through the Assessment List for Trustworthy AI (ALTAI), the AI principles are

⁵ Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, Luetge C, Madelin R, Pagallo U, Rossi F, Schafer B, Valcke P, Vayena E. AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines* 28, 689-707. 2018. doi: 10.1007/s11023-018-9482-5.

⁶ European Commission, Directorate-General for Communications Networks, Content and Technology. Ethics guidelines for trustworthy AI. Publications Office. 2019, <https://data.europa.eu/doi/10.2759/346720>.

⁷ D. Leslie, Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector, 2019, The Alan Turing Institute. <https://doi.org/10.5281/zenodo.3240529>

⁸ UNESCO World Commission on the Ethics of Scientific Knowledge and Technology, Preliminary study on the Ethics of Artificial Intelligence, 2019, <https://unesdoc.unesco.org/ark:/48223/pf0000367823>.

⁹ European Commission, Directorate-General for Communications Networks, Content and Technology. Ethics guidelines for trustworthy AI. Publications Office. 2019, <https://data.europa.eu/doi/10.2759/346720>.

¹⁰ European commission, Directorate-General Research & Innovation, Ethics By Design and Ethics of Use Approaches for Artificial Intelligence, 25 November 2021, https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf.

¹¹ OECD. Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. 2022, available online at <https://legalinstruments.oecd.org/api/print?ids=648&lang=en>.

¹² Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts COM/2021/206 Final, [https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=en&reference=2021/0106\(COD\)](https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=en&reference=2021/0106(COD)) (for subsequent versions of the document).

translated into an accessible and dynamic checklist that guides developers and deployers of AI in implementing such principles in practice, indicating a set of concrete steps for self-assessment.

On the other hand, the CEPEJ's Ethical Charter represents, so far, the only attempt to provide an overarching ethical framework for the development and use of AI system in the judiciary, which is the specific focus of the ADELE Project.

We did not consider legal compliance under the proposed AI Regulation as its approval is pending and amendments may still be introduced in the final text.

In the remainder of the section, we shall present the two above-mentioned documents to give a flavor of ethical challenges pertaining to the context of AI applications, possibly related to the field of justice.

1) High-Level Expert Group's Ethics Guidelines for Trustworthy Artificial Intelligence

The Guidelines by HLEG are general guidelines applicable to all AI systems. They affirm the s.c. "trustworthy paradigm", namely that AI should be developed (1) lawfully, complying with all applicable laws and regulations (2) ethically, ensuring adherence to ethical principles and values and (3) that it should be robust, both from a technical and social perspective since, even with good intentions, AI systems can cause unintentional harm.

The Guidelines requires that AI systems should align with ethical norms, in particular with the following principles:

1. **Respect for human autonomy**, according to which humans interacting with AI systems must be able to keep full and effective self-determination over themselves and be able to partake in the democratic process. AI systems should not "unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans, but rather, argument, complement and empower human cognitive, social and cultural skills, leave opportunity for human choice and securing human oversight over work processes" in AI systems, e.g., support humans in the work environment and support the creation of meaningful work.
2. **Prevention of harm**, according to which AI systems should neither cause nor exacerbate harm or otherwise adversely affect human beings. AI systems should "protect human dignity, mental and physical integrity, be technically robust and assure they are not open to malicious use". For example, they should be supervised to avoid such systems exacerbating adverse impacts due to information asymmetries or unequal balance of power.
3. **Fairness**, according to which the development, deployment and use of AI systems must be fair, both from a substantive and procedural point of view. Fairness is closely related to the rights to non-discrimination, solidarity and justice. Despite the varying interpretations of fairness, the European Commission advocates for having both: (a) a substantive dimension of fairness that "commits to ensure equal and just distribution of benefits and costs, commits to free from unfair bias, discrimination and stigmatization, implies respecting the principle of proportionality between means and ends and a careful balancing of competing interests and objectives", and (b) a procedural dimension allowing to "contest and seek redress against decisions taken by AI systems or who operates them". To achieve this, focus is placed on the identifiability of the entity responsible for the decision and the explainability (also referred to as "interpretability") of the decision-making processes.

4. **Explicability**, according to which the processes engaged with by AI systems need to be transparent and their capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected. Depending on the algorithms used, a model generating a particular output (decision) may not always be fully explained – in such cases, “other explicability measures (e.g.. traceability, auditability and transparent communication on system capabilities) may be required, provided that the system as a whole respects fundamental rights. The degree to which explicability is needed is highly dependent on the context and the severity of the consequences if that output is erroneous or otherwise inaccurate.”

The four principles are further translated in seven requirements, which must be implemented by both the developer and the deployer to reach a trustworthy AI system:

1. **Human agency and oversight**, including the respect of Including fundamental rights, human agency and human oversight;
2. **Technical robustness and safety**, including resilience to attack and security, fall back plan and general safety measures, as well as accuracy, reliability and reproducibility measures;
3. **Privacy and data governance**, which include the respect for privacy and data protection and the overall data governance plan, including data quality, integrity and access;
4. **Transparency**, including traceability, explainability of the technical process and decision possibly taken, and relevant communication to all relevant stakeholder as to the nature and capacity of AI systems;
5. **Diversity, non-discrimination and fairness**, which include avoidance of unfair bias, accessibility and universal design measures and participation of all relevant stakeholders;
6. **Societal and environmental well-being**, including a sustainable and environmentally friendly AI, controlling its social impact, and upholding societal progress and democracy;
7. **Accountability**, including auditability, detection of risks, minimization and reporting of harms, and accountability in trade-off making and redress measures.

2) European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment

The Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment issued by the CEPEJ in 2018 is intended for public and private stakeholders responsible for the design and deployment of AI tools and services that involve the processing of judicial decisions and data. It also concerns public decision-makers in charge of the legislative or regulatory framework, of the development, audit or use of such tools and services.

According to the Charter, the use of such tools and services in judicial systems seeks to improve the efficiency and quality of justice and should be encouraged. However, their development should be carried out responsibly, with due regard for the fundamental rights of individuals as set forth in the European Convention on Human Rights and the Convention on the Protection of Personal Data, and in compliance with other fundamental principles set out below:

1. **Respect for fundamental rights:** ensure that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights. It is explicitly mentioned that when AI tools are used as a tool to assist in judicial decision-making, it is essential to ensure that they do not undermine the guarantees of the right of access to the judge and the right to a fair trial. Further, such tools should also be used with due respect to the principles of the rule of law and judges' independence in their decision-making process. Mechanisms for redress and recourse should be in place to address any violations of fundamental rights caused by AI tools. This may involve establishing channels for reporting issues, offering the possibility of human review when contested decisions are made, and providing remedies for individuals adversely affected by the AI system's actions.
2. **Non-discrimination:** specifically prevent the development or intensification of any discrimination between individuals or groups of individuals. This principle reflects specifically on the grouping or classifying data relating to individuals or groups of individuals, and the processing – directly or indirectly – of “sensitive” data¹³. It underscores the importance of preventing the development or exacerbation of discriminatory practices towards individuals or groups. AI systems should be designed and implemented in a manner that upholds fairness, impartiality, and equitable treatment for all.
3. **Quality and security:** with regard to the processing of judicial decisions and data, use certified sources and intangible data with models elaborated in a multi-disciplinary manner, in a secure technological environment. This principle emphasizes the need to maintain high standards in the sources of information and data used, as well as in the technological environment where AI models operate. When processing judicial decisions and data, it is crucial to rely on certified and authoritative sources. Certified data can be validated and cross-checked to prevent the propagation of misinformation or biased content. In addition, developing AI models for processing judicial decisions requires collaboration between experts from different disciplines. Legal experts, data scientists, ethicists, and other relevant professionals should work together to ensure that the AI system's design aligns with legal principles, ethical considerations, and technical feasibility.
4. **Transparency, impartiality and fairness:** make data processing methods accessible and understandable and authorise external audits. This transparency is essential for judges, lawyers, and individuals involved, in order to understand the reasoning behind the outcomes. The design of the AI system should seek a balance between the intellectual property of certain processing methods and the need for transparency (both in terms of technical and non-technical transparency¹⁴), impartiality (absence of bias), fairness and intellectual integrity (prioritizing the interests of justice). AI models used in the judicial context should undergo continuous evaluation and improvement. Regular assessments help identify any shortcomings or biases and allow for timely adjustments to enhance the system's fairness and effectiveness.
5. **User control:** preclude a prescriptive approach and ensure that users are informed actors and in control of the choices made. This principle incorporates several aspects of user control: user autonomy, human-in-the-loop approach (professionals in the justice system should, at any

¹³ This could include alleged racial or ethnic origin, socio-economic background, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, health-related data or data concerning sexual life or sexual orientation.

¹⁴ The functioning of the AI system should be explained in clear and familiar language (e.g., to describe how results are produced) by communicating the nature of the services offered, the tools that have been developed, performance and the risks of error, and so on.

moment, be able to review judicial decisions and the data used to produce a result and continue not to be necessarily bound by it in the light of the specific features of that particular case), and communicating about the system and its features in a clear and understandable language of any prior processing of a case by artificial intelligence before or during a judicial process (thus, having the right to object).

3. Compliance measures

Based on the abovementioned ethical guidelines, the ADELE team has implemented a number of compliance measures, related to:

1) Human agency and user control

The ADELE Pilot Tool is designed in a way to enable human oversight (human-in-the-loop, human-on-the-loop, human-in-command), traceability and auditability. It is not developed as a stand-alone tool for decision-making but as a support instrument for selected and domain-specific insights to judges and other legal professionals, and its use is explicitly triggered by the interested professionals for exploratory purposes.

Any outcome of the “Outcome prediction” Module¹⁵ is to be intentionally triggered by a professional, for a particular domain and language, by filling in a form and introducing the requests, claims and related arguments they are interested in by themselves. Once the result is shown, it is again left to the professional how to use this information. Moreover, the professional is explicitly warned that the result was obtained using AI algorithms on the number of decisions (including in what domain and language) taken into consideration to train the model, and the fact that the prediction of the outcome of the case was not binding and was not intended to replace professional legal advice.

The user manual¹⁶ provides easy and comprehensible information on the use of the pilot tool’s different functionalities, as well as the assumption behind machine learning models and related accuracy.

Further, the guidelines¹⁷ used to annotate case law for machine learning have also been published and are available on the webpage of the platform. While they are prepared in a more technical language (considering the nature of the document), they have already been “tested” by legal professionals involved in the project, and they could be followed without great difficulties by a non-technical expert after a careful review.

Testing and training events have been organized at the national level (in both Bulgaria and Italy) to educate judges and other legal professionals potentially using the ADELE Pilot Tool to understand the various functionalities of the tool and the potential effects of its use. The experience in the training events has shown no specific comments were made regarding the use of the tool concerning human agency and user control, and the matter was considered clear by the audiences. However, discussions

¹⁵ For example, using the following link <https://adele-tool.eu/it/2/1/Doc/OutcomePrediction> one explores outcomes predictions based on 230 decisions of Italian courts on claims for infringements of rights on trademarks and patents.

¹⁶ <https://adele-tool.eu/api/data/userManual>, <https://site.unibo.it/adele/en/publications/tools>

¹⁷ <https://site.unibo.it/adele/en/publications/tools>.

were raised regarding the numbers of annotated decisions and whether they could be considered sufficient for accuracy purposes, which showed sufficient understanding of the features presented and the critical thinking process of the professionals.

2) Respect of fundamental rights in judicial proceedings

The processing of judicial decisions and data must serve clear purposes, in full compliance with the fundamental rights guaranteed by the European Convention on Human Rights (ECHR) and the applicable data protection legislation.

The clear purpose of the ADELE Pilot Tool has been extensively explained on the project website, tool's platform, and all related documents and events (as explained previously).

Protection of personal data, in particular, has also been discussed in detail in Deliverable D2.3 (Final annotated corpus). Thus, we will avoid further repetition here. The principles of privacy by design and by default have been applied to secure compliance with the principles of lawfulness, data minimisation, transparency, and fairness.

As to the respect of the fundamental rights beyond data protection, it must be stated that the ADELE Pilot Tool, including the "Outcome prediction" Module, has been designed in such a way as to comply with the European Convention on Human Rights (ECHR) from the very beginning. However, in the context of the judicial proceedings, the right of access to the judge (court), the right to a fair trial, the principles of the rule of law and judges' independence in their decision-making process require further ex-post investigation.

Article 6(1) ECHR lists several elements comprising the fair administration of justice. While the paragraph describes – to some extent – the general characteristics of judicial institutions and outlines the broad parameters by which the fairness of a proceeding can be judged, what is imperative is that the individual has the opportunity to have her/his case heard in the first place, and subsequently it is important that the court's decision becomes operative as soon as possible. With regard to access to court, there are several important principles: the state cannot restrict or eliminate judicial review in certain fields or for certain classes of individuals, access to a judicial forum must be substantive, not just formal, and the third aspect concerns the actual implementation of the right.

In the determination of one's civil rights and obligations or of any criminal charge against him/her, everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law. This encompasses several aspects: the hearing is held within a reasonable time, one is heard by an independent and impartial decision-maker, one is given all the relevant information, the hearing is open to the public (although the press and public can be excluded for highly sensitive cases), one is allowed representation and an interpreter where appropriate, and the hearing is followed by a public decision. One also has the right to an explanation of how the court or decision-making authority reached its decision.

The relevance of the rule of law is demonstrated by application of the following principles in practice: the law is applied equally and fairly, so that no one is above the law; the separation of powers between the legislature, the executive and the judiciary; the judicial system is independent and impartial with open justice; the law is made by representatives of the people in an open and transparent way; the

law is capable of being known by everyone, so that everyone can comply; people can only be punished in accordance with the law; no one is subject adversely to a retrospective change of the law or prosecuted, for any offence not known to the law when committed; government agencies are to act as model litigants; a fair and prompt trial is in place; all people are presumed to be innocent until proven otherwise and are entitled to remain silent and are not required to incriminate themselves; and last but not least, the law and its administration is subject to open and free criticism by the people, who may assemble without fear.

However, none of these aspects is threatened by the ADELE Pilot Tool. The use of the tool is voluntary. It has an exploratory purpose rather than expert decision-making support and its use could be contested at any time, if applied in practice. Also, in case of the latter, the court and the judge still bear the responsibility to assess whether the tool fit for the purpose of their use, and that they have taken all precocious measures to avoid disrespecting the fundamental rights of the party in question. Further, the system's results have been tested for reproducibility (i.e., the extent to which a tool is capable of producing the same result when used repeatedly in the same circumstances), and it has proven its reliability. Therefore, there are no indications of unequal treatment.

Judges' independence refers to the judiciary remaining impartial to decide matters before them, on the basis of facts and in accordance with the law, without any restrictions, improper influences, inducements, pressures, threats or interferences, direct or indirect, from any quarter or for any reason. In the context of impartiality, the ADELE team have taken all measures to ensure high levels of awareness among the judges towards the nature of the ADELE Pilot Tool, its purposes and the fact that its model has been trained to a limited number of decisions in only two legal domains. However, proper usage of the tool requires not only awareness of its limitations but also sufficient knowledge and understanding of how the AI techniques work, and the ADELE team has no way to ensure such knowledge and understanding in all possible users beyond the information and documents already provided. It is then on the conscience of the judge to decide whether the use of the tool could support an impartial and based-on-facts decision.

3) Fairness and non-discrimination

Principles of discriminatory non-harm and fairness entail that both designers and users ensure that the AI systems they are developing and deploying: are trained and tested on properly representative, relevant, accurate, and generalisable datasets (data fairness); have model architectures that do not include target variables, features, processes, or analytical structures (correlations, interactions, and inferences) which are unreasonable, morally objectionable, or unjustifiable (design fairness); do not have discriminatory or inequitable impacts on the lives of the people they affect (outcome fairness); and are deployed by users sufficiently trained to implement them responsibly and without bias (implementation fairness).

The ADELE Pilot Tool is designed in a way to avoid bias in both input data and algorithm design. In particular, the legal domains chosen for piloting the ADELE Pilot Tool (VAT and trademarks and patents) do not reflect on special categories of personal data (racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data

concerning a natural person's sex life or sexual orientation) and no such information should be considered even slightly relevant for the selected cases. Nevertheless, the ADELE team has ensured training data used to train machine learning models have been properly anonymised and abide all relevant statistical properties in that sense. The rigorous testing process have not revealed any tendencies towards discrimination, stigmatisation or any other adverse effects on the individuals and groups referred to in the cases, part of the training datasets. Further, considering the input data (requests, claims, arguments) related to the selected legal domains, there is no indication any sensitive data of the user (judge, legal professional) is taken into consideration while forming the output.

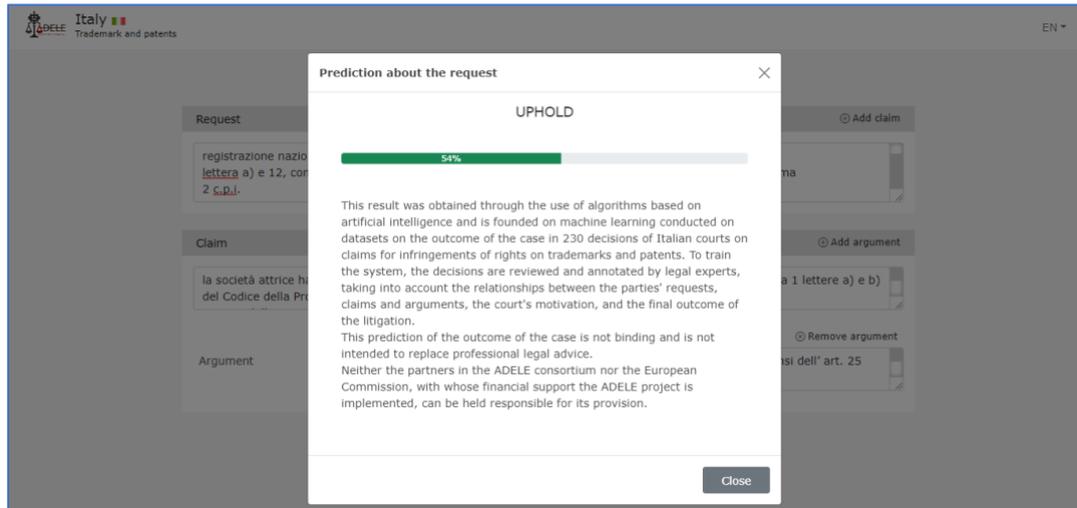
As mentioned, the ADELE team have put special efforts into preparing information and documents to make judges and other legal professionals aware of the ADELE model's limitations imposed by the limited training sets used while providing them with substantive material to understand the specifics of the model and how it was trained.

4) Transparency

ADELE Pilot Tool users are made aware they are interacting with an AI system and are informed, in plain and clear language about its abilities, limitations, risks and benefits. Whenever possible, an explanation of how decisions are made by the model is provided to the users. The ADELE Pilot Tool webpage clearly indicates that this is a piloting solution, designed to foster the scientific debate on the use of AI in the legal field. Also, the way the features are divided (based on a selection of territorial scope and legal domain) shows the purposeful limitations introduced by the use of AI techniques under the project.

Due to time and effort constraints, the current version of the Pilot Tool is not fully explainable, especially with regard to the "Outcome prediction" Module, though some experiments on feature analysis have been carried out. Nevertheless, users interacting with the "Outcome prediction" Module are informed about that through the following message (in the example of one exploring outcome predictions based on 230 decisions of Italian courts on claims for infringements of rights on trademarks and patents):

"This result was obtained through the use of algorithms based on artificial intelligence and is founded on machine learning conducted on datasets on the outcome of the case in 230 decisions of Italian courts on claims for infringements of rights on trademarks and patents. To train the system, the decisions are reviewed and annotated by legal experts, taking into account the relationships between the parties' requests, claims and arguments, the court's motivation, and the final outcome of the litigation. This prediction of the outcome of the case is not binding and is not intended to replace professional legal advice. Neither the partners in the ADELE consortium nor the European Commission, with whose financial support the ADELE project is implemented, can be held responsible for its provision."



5) Technical safety

In the context of the ADELE Pilot Tool, technical safety refers to both information security and the usage of certified sources and intangible data with models conceived in a multidisciplinary manner.

On the one hand, the ADELE team incorporates and draws widely on the expertise of legal professionals with experience and knowledge in the legal domains chosen for the purpose of piloting its tool. The ADELE team itself is a multidisciplinary one with all relevant expertise being involved in producing the functional model capitalizing on this very same multidisciplinary. On the other hand, all judicial decisions on which the ADELE model is trained come from certified sources and have not been modified but enhanced during the annotation process in a purely technical manner to allow for the training of the learning mechanism to take place. The annotation process is fully traceable to ensure that no modification has occurred to alter the content or meaning of the decision being processed.

To ensure all data, models and algorithms are stored and executed in a secure environment (thus, ensuring system integrity and intangibility), applicable technical standards and information security best practices have been applied. No negative impact or vulnerabilities have been examined.

6) Accountability

The ADELE Pilot Tool is designed by a diverse team of experts familiar with the legal and ethical framework regarding the use of artificial intelligence techniques. Fundamental rights protection, and personal data protection in particular, have been discussed at various stages and shaped the final set of features chosen for the tool. In addition, the project implementation structure incorporates a proper level of transparency and oversight (both horizontally and vertically) and all team members are aware of the tool's characteristics, features and input/output specifics and the respective legal and ethical restrictions therein. Activity monitoring protocols have been in place to enable end-to-end oversight and review.

Maintaining a culture of responsibility, the ADELE Consortium has set up an internal ethics board with appropriate and diverse expertise (artificial technology techniques, judicial decision-making, and ethics of new and emerging technologies) to ensure better accountability. In case the ADELE

Consortium is alerted of any potential violation of the ethical principles explored in the present document, it will be examined by the ethics board to ensure the ethical, fair, and safe application of AI techniques.

4. Conclusions

The ethics review described herein has shown the ADELE team has taken sufficient measures to explain to the end-users of the ADELE Pilot Tool (judges and other legal professionals) their interaction with a system integrating AI techniques, as well as its abilities, limitations, potential risks and benefits, and the manner in which the outcomes are produced. It has taken specific measures to avoid bias in input data and algorithmic design. The team has also ensured that the ADELE Pilot Tool complies with the fundamental rights protection framework and that any potential ethics risks are mitigated by a proper accountability structure.

Considering the ADELE Pilot Tool is not intended and designed to interact and/or replace human decision-making processes, the transparency measures taken are deemed sufficient to avoid influencing the end-users in such a way to lead to negative impact on individuals and groups and/or potential infringement of fundamental rights of the parties in judicial proceedings.