# A Risk-based Approach to Trustworthy AI Systems for Judicial Procedures

Majid Mollaeefar[1,*], Eleonora Marchesini[1], Roberto Carbone[1] and Silvio Ranise[1,2]

[1]Fondazione Bruno Kessler, Center for Cybersecurity, Trento, Italy
[2]Department of Mathematics, University of Trento, Italy

**Abstract**

In the rapidly evolving landscape of Artificial Intelligence (AI), ensuring the trustworthiness of AI tools deployed in sensitive use cases, such as judicial or healthcare processes, is paramount. The management of AI risks in judicial systems necessitates a holistic approach that includes various elements, such as technical, ethical considerations, and legal responsibilities. This approach should not only involve the application of risk management frameworks and regulations but also focus on the education and training of legal professionals. For this, we propose a risk-based approach designed to evaluate and mitigate potential risks associated with AI applications in judicial settings. Our approach is a semi-automated process that integrates both user (i.e., judge) feedback and technical insights to assess the AI tool's alignment with Trustworthy AI principles.

**Keywords**

Judicial AI, Risk-aware, Trustworthy AI, Trustworthiness Risk Assessment.

## 1. Introduction

In recent years, the adoption of Artificial Intelligence (AI) technologies has surged across various industries and domains. AI systems now play a pivotal role in making critical decisions, automating tasks, and augmenting human capabilities. However, with the expanding influence and complexity of AI, it is crucial to ensure the development and deployment of Trustworthy AI (TAI) systems. TAI encompasses the creation and implementation of AI technologies adhering to a set of principles that promote transparency, fairness, accountability, and robustness. By designing TAI systems, the aim is to inspire trust among users, stakeholders, and society as a whole where these systems must operate reliably, ethically, and in a manner that respects fundamental rights and values. The significance of TAI cannot be overstated, as it has the potential to address pressing concerns that arise from increasing reliance on AI systems. Some notable reasons why it is critical for AI systems to be designed with trustworthiness in mind including the following three; First, TAI cultivates user confidence and trust by ensuring that personal data is handled responsibly, decisions made by AI systems are fair and unbiased, and privacy is protected. This is critical for building user confidence and trust in AI systems. The authors in [1] discuss the theoretical framework of AI trustworthiness, including aspects of privacy preservation and fairness, which are key to fostering user trust. Second, TAI bolsters the accountability and explainability of AI systems. As these systems become integral to decision-making processes, it is essential to comprehend how they reach their conclusions or recommendations. TAI increases transparency and offers mechanisms for interpreting the rationale behind AI-generated decisions, allowing users and stakeholders to hold systems accountable. Cobianchi et al. [2] emphasize the importance of accountability, technical robustness, and transparency in AI applications in surgery, which can be extended to other domains. Third, TAI aids in mitigating risks associated with AI technologies. If developed or deployed irresponsibly, AI systems can introduce numerous risks, including privacy breaches, biased decision-making, safety concerns, and the perpetuation of social inequalities. Addressing these risks is vital to protect individuals, organizations, and society from potential harm and adverse consequences.

The AI Act draft proposal for a Regulation[1] of the European Parliament and of the Council laying down harmonized rules on AI represents the first attempt to enact a horizontal AI regulation. This proposed legal framework, focusing specifically on the use of AI systems, advocates for a technology-neutral definition of AI systems in EU legislation. It emphasizes a risk-based approach where AI systems are classified with varying obligations proportional to their level of risk. The AI Act categorizes risks into four levels: minimal, limited, high, and unacceptable (i.e., the latter are not permitted to be sold on the EU market). It focuses on high-risk AI applications (HRAI) by setting specific requirements and obligations for both users and providers of these applications. This includes a conformity assessment before market place-

*Ital-IA 2024: 4th National Conference on Artificial Intelligence, organized by CINI, May 29-30, 2024, Naples, Italy*

*Corresponding author.
mmollaeefar@fbk.eu (M. Mollaeefar); emarchesini@fbk.eu (E. Marchesini); carbone@fbk.eu (R. Carbone); ranise@fbk.eu (S. Ranise)



[1]https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206

ment or service commencement, enforcement measures post-market placement, and a governance structure at both European and national levels. The aim is to ensure that obligations are aligned with the associated risk level of each AI system.

One of the areas where AI holds a sensible impact is in the legal context, where for instance judges can benefit from the presence of automated decision-making in judicial proceedings [3, 4], potentially reducing the effort required to search through documents, seek out relevant legal provisions, or support them in complex cases where the human capacity to detect patterns is limited [5]. AI tools like ChatGPT, while useful, present several limitations in legal contexts. They may produce inaccurate information, as demonstrated in cases like Roberto Mata vs Avianca[2], where reliance on ChatGPT led to legal issues due to the citation of non-existent cases. This stresses the necessity for legal professionals, particularly judges, to be acutely aware of the risks associated with their use of HRAI Systems. In this paper, we introduce a risk-based approach designed to evaluate and mitigate potential risks associated with the trustworthiness of AI applications in judicial settings.

# 2. Background

Below we introduce background information to better perceive the approach.

## 2.1. AI Algorithms

In the realm of AI, the development of algorithms falls into two primary views: traditional and modern. The traditional approach involves human-created models for specific problems or computations, where a limited set of features and a fixed sequence of instructions are employed. This method, exemplified by classical planning in autonomous systems, relies on symbolic representations and a predefined set of rules, necessitating heuristics to navigate the vast potential state spaces. Despite its rigidity, this approach allows for the construction of algorithms that are easily understood and verified by humans. Conversely, the modern perspective, dominated by Machine Learning (ML), leverages large datasets to generate rules for problem-solving. Through processes like training and deployment, algorithms are formulated to classify or interpret data, such as classifying images of dogs and cats. The ML-based methods benefit from the ability to tackle complex problems without extensive human ingenuity, employing powerful optimization techniques. However, it faces challenges such as potential imprecision, bias in training data, and the complexity of the resulting algorithms making them difficult for humans to comprehend. Strategies to mitigate these issues include performance monitoring, dataset filtering, and developing techniques for better human understanding of ML-generated algorithms. The choice between traditional and modern methods depends on the specific application's needs, including considerations of security and trustworthiness. An effective risk analysis is crucial in determining the suitability of an AI-produced algorithm for a given scenario.

## 2.2. Trustworthy AI

Trustworthiness is a prerequisite for people and societies to develop, deploy and use AI systems. Without AI systems—and the human beings behind them—being demonstrably worthy of trust, unwanted consequences may ensue, and their uptake might be hindered, preventing the realization of the potentially vast social and economic benefits that they can bring [6]. In the past few decades, the success of ML has primarily been evaluated based on its quantitative accuracy, which has made training AI models much more manageable. Predictive accuracy has also become the standard measure for determining the superiority of an AI product. However, with the widespread use of AI, the limitations of using accuracy as the sole measurement have become apparent, as new challenges have arisen, such as malicious attacks and the misuse of AI. To address these challenges, the AI community has recognized that factors beyond accuracy need to be considered and improved when building an AI system. Recently, a number of enterprises, academia, public sectors, and organizations have identified principles of AI trustworthiness that go beyond accuracy-based measurements [7]. According to [8], the current degree of trustworthiness of an AI system is dependent on how the user perceives its technical characteristics. Various organizations, including the G20, the EU Parliament, the General Partnership on AI (GPAI), and the Organisation for Economic Co-operation and Development[3] (OECD) have proposed different principles for ensuring trustworthiness in AI systems [9]. The OECD, for instance, has put forward a set of five principles aimed at promoting TAI: (i) inclusive growth, sustainable development and well-being, (ii) human-centered values and fairness, (iii) transparency and explainability, (iv) robustness, security and safety, and (v) accountability. The use of AI is intended to promote human good and well-being, and as such, it should not cause any harm. AI systems must be characterized by fairness, accuracy, and reliability, and should not be discriminatory. To be considered trustworthy, AI systems must be transparent and explainable, meaning they should have the necessary capabilities, functions, and features to achieve user goals, with their algorithms being easily understood by users. Additionally, AI systems must be resilient to threats that may try to exploit their normal behaviors and turn them into harmful ones. In the literature, additional principles have

[2] https://law.justia.com/cases/federal/district-courts/new-york/nysdce/1:2022cv01461/575368/54/

[3] https://oecd.ai/

been proposed such as accuracy [10], acceptance [11], predictability and performance [12]. The AI HLEG [6], has focused on the concept of TAI, offering guidance in the form of a framework and identifying seven key ethical and technical requirements.

## 3. Our View on Trustworthiness

In our analysis of the literature on finding principles of trustworthiness in AI, the commonly agreed-upon principles are *accuracy*, *robustness*, *privacy*, *explainability*, *accountability*, and *fairness*. While these six principles are widely acknowledged in the literature, there are additional considerations that can be incorporated within them. For instance, the concept of "human in the loop" can be viewed as an aspect of fairness. We differentiate between properties and principles. While both concepts are related and work together to ensure the overall trustworthiness of AI systems, they represent different aspects of the trustworthiness framework. Properties refer to specific characteristics or attributes of an AI system that contribute to ensure a principle. For instance, integrity, reliability, and data validity can be considered as properties relevant to the accuracy principle; *Integrity* refers to the quality of an AI system being honest, consistent, and maintaining the integrity of the data and algorithms it operates on. It ensures that the AI system is resistant to unauthorized modifications or tampering. *Reliability*, focuses on the consistency and dependability of an AI system's performance. A reliable AI system consistently produces accurate results over time and under different conditions. *Data validity* refers to the quality and correctness of the data used by an AI system to generate outputs. Valid data ensures that the information processed by the AI system is accurate, relevant, and representative of the problem domain. On the other hand, principles represent high-level guidelines or concepts that guide the development and deployment of TAI systems. The relationship between properties and principles lies in how properties contribute to fulfilling the principles. Figure 1 depicts the relationship between properties and six essential principles for TAI, categorized into either technical, ethical, or both. *Accuracy* and *robustness* serve as technical principles, whereas *fairness* and *accountability* fall within the ethical domain. Located in the center of the figure, *privacy*, and *explainability* are unique principles that encompass both the technical and ethical facets.

### 3.1. AI Algorithms & Trustworthiness

Trustworthiness in AI is a multifaceted concept, often seen as a relationship between two entities—the AI system and its user. The trustworthiness of an AI system is largely dependent on how it is perceived by the user in terms of its technical characteristics. This perception is influenced by various factors, including the type of AI model, its application context, and the underlying algorithms [13]. Different AI models exhibit variability in how they align with TAI principles. This variation stems from the inherent differences in model structures, training methods, data used, and their intended applications. For example, a model designed for healthcare decision support may prioritize accuracy and privacy, while one for autonomous vehicles might focus more on safety and robustness. The data used to train AI models significantly affects their trustworthiness. A model trained on limited or biased data may exhibit lower trustworthiness due to its potential to generate skewed or unfair results. Additionally, the type of algorithm—whether it is rule-based or learning-based—plays a crucial role in determining the model's reliability, fairness, and transparency [13].

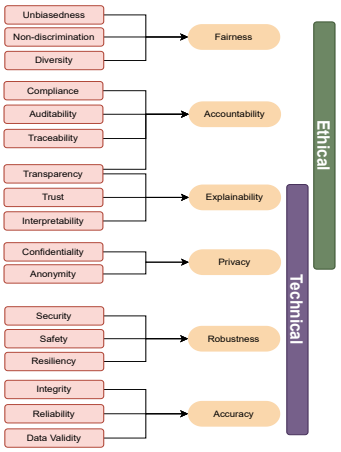


**Figure 1:** TAI principles and properties relationship.

### 3.2. Algorithm-based Trustworthiness

The relationship between algorithms and TAI principles is a critical aspect of responsible AI development and deployment. TAI principles serve as benchmarks against which the performance and ethical considerations of algorithms can be evaluated. Each algorithm has its own set of advantages and limitations that align or conflict with these principles, making it essential to investigate their compatibility in specific use cases. Since each algorithm has a distinct set of characteristics, their compatibility with TAI principles can differ significantly; in other words, they have different compliance levels. To define Algorithm-based Trustworthiness (ABT) levels, it is essential to consider both the inherent characteristics of each algorithm and the specific attributes related to each AI principle. We define the following qualitative levels for this assessment; **High**: The algorithm inherently aligns with the AI principle in question, requiring minimal or no additional measures to ensure compliance. **Moderate**: While the algorithm generally aligns with the principle, additional safeguards or contextual consid-

erations may be necessary. **Low**: The algorithm poses challenges or risks that make it difficult to align with the AI principle, and significant adjustments or limitations would be required for compliance. To conduct a comparison between rule-based and ML-based AI algorithms, we need to consider some assumptions such as consistency of environment (i.e., static or dynamic), the complexity of problems, availability and quality of data, risk of bias, need for transparency, and explainability. With these considerations, in our judicial case, we take these assumptions; (i) the operational environment for the AI system is dynamic, (ii) the complexity of the problem can be considered as high, (iii) the high quality of datasets are available, free of bias and sensitive personal information, and (iv) the explanation of the decisions is required. With these considerations, in the following, we qualitatively evaluate the compatibility of the two distinct types of algorithms with TAI principles.

#### 3.2.1. Rule-based AI

These AI systems are perfectly suited to applications that require small amounts of data and simple, straightforward rules. These algorithms exhibit high accuracy due to deterministic outcomes from well-defined rules. However, since the assumption of the operational environment is dynamic and the problem is complex, we consider a moderate level for the *accuracy* principle. These algorithms can be very robust if the rules are well-crafted to handle various edge cases. But they may falter in scenarios not covered by the existing rules, therefore, their *robustness* can also be considered *moderate*. These algorithms stand out for their high *explainability* and *accountability*, as their rule-based nature makes them transparent and easy to understand, even for non-experts.

#### 3.2.2. ML-based AI

These AI systems, particularly suited for environments with abundant data, vary in their alignment with TAI principles. For the sake of simplicity, we focus only on four key supervised ML models; Linear Regression (LR), Decision Trees (DT), Support Vector Machines (SVM), and Neural Networks (NNs). LR is chosen for its fundamental approach to data modeling. DTs offer a more intricate decision-making structure. SVMs are known for their efficiency in high-dimensional spaces, while NNs, especially in deep learning, handle complex tasks like image and language processing. These models collectively represent the diverse capabilities of ML and provide insights into their trustworthiness in dynamic, data-intensive scenarios. For *accuracy* and *explainability* principles, there is a notable trade-off observed across the algorithms. In the literature [14, 15], there has been a comprehensive comparison of different ML models in terms of their accuracy and explainability level. The LR and DT algorithms, while offering high levels of explainability due to their transparent nature, may not achieve the same level of accuracy in complex scenarios as their more sophisticated counterparts. On the other hand, SVMs and neural networks, especially in their advanced forms, are capable of handling complex, high-dimensional data with greater accuracy but often sacrifice explainability, presenting a challenge in understanding the rationale behind their decisions. When it comes to *robustness*, SVMs are distinguished by their high resilience, particularly against adversarial attacks, thanks to their strong generalization capabilities. NNs, despite their adeptness at complex pattern recognition, exhibit moderate to low robustness and are vulnerable to adversarial examples, requiring specialized methods like adversarial training to enhance their robustness. DTs offer a moderate level of robustness, valued more for their interpretability than their resistance to adversarial examples, while LR models are less robust, particularly in complex datasets and adversarial environments. In terms of *accountability*, LR models excel due to their straightforward and transparent nature, which makes tracing decisions back to specific data points relatively easy. DTs also score highly in this regard, due to their clear decision-making paths. SVMs, particularly with non-linear kernels, present a more complex picture, offering moderate to low accountability due to the intricacies involved in their decision-making processes. NNs are at the lower end of the spectrum in terms of *accountability*, often described as "black boxes" due to their complex, layered structures, although efforts like layer-wise relevance propagation (LRP) and SHAP[4] values are employed to enhance their interpretability. The aspects of *fairness* and *privacy* are also pivotal in evaluating the TAI alignment of ML algorithms. The fairness of algorithms such as LR, DTs, SVMs, and NNs is predominantly governed by the nature of their training data. Since these algorithms inherently lack bias, any unfairness in decision-making largely stems from biases present in the training data. This reality highlights the importance of precise data collection and processing, ensuring that the data is representative and free of biases to maintain fairness in the outcomes. Alongside fairness, privacy considerations in these algorithms are crucial, yet they are not intrinsic to the algorithms themselves. Instead, privacy risks are closely tied to how the data is handled. Ensuring the privacy and security of data, especially sensitive personal information, is vital, regardless of the algorithm in use. Effective data handling practices, including anonymization and secure storage, play a critical role in mitigating privacy risks in machine learning applications. Therefore, in both fairness and privacy, the emphasis shifts from the algorithmic design to the careful management of the data they process. In Table 1, we summarized the ABT levels for rule-based and ML-based algorithms. This

[4] https://github.com/shap/shap

**Table 1**
Qualitative comparison between the algorithms and their alignment with TAI principles. Legend; **L**ow, **M**oderate, **H**igh

| TAI Principles | Rule-based | ML-based (Supervised) | | | |
|---|---|---|---|---|---|
| | | LR | DT | SVM | NNs |
| Accuracy | M | L | H | H | H |
| Robustness | M | L | H | M | M |
| Accountability | H | H | M | L | L |
| Explainability | H | H | M | L | L |
| Privacy | Depends on data handling, not inherent to the model. | | | | |
| Fairness | Depends on the data pipeline. | | | | |

comparison, which provides a framework to gauge how various algorithms align with TAI principles, supports the risk assessment process effectively. In the next section, we will propose a risk-based approach, where these comparative insights become a vital factor in evaluating AI trustworthiness and assessing risk levels.

# 4. The Risk-based Approach

The primary goal of this approach is to support judges and legal practitioners with a set of best practices when utilizing AI tools in their judicial work. This includes providing them with a clear understanding of the potential risks associated with these tools and offering actionable suggestions to mitigate these risks, ensuring responsible and informed use of AI in legal settings. The approach is a semi-automated process that requires user interaction at the beginning of the approach to collect useful information about the AI tool. This approach assesses risks associated with the use of AI tools, focusing on their alignment with TAI principles and their role in legal contexts. Before diving into the approach, we consider some assumptions; (i) the user has some experience using the AI tool, (ii) the user does not know anything about the technical details behind the AI tool, (iii) the user knows only about the required input and output. Typically *risk* defines as a function of two values **L**ikelihood and **I**mpact (i.e., Risk = $f$(**L**,**I**)). Similarly, we formulate the likelihood as function of two values which are ABT and Control effectiveness (CE), where the *ABT* refers to the degree to which the AI tool's algorithm aligns with TAI principles. It assesses whether the algorithmic design and functionality inherently support or conflict with these principles. For instance, the tool utilized with deep neural networks has a high level of accuracy in prediction while their "black-box" nature makes them less explainable (see Table 1). Instead, the *CE* represents the effectiveness of implemented controls in mitigating risks associated with the AI tool. For example, strict access controls and logging mechanisms increase confidentiality mitigate the risk to the privacy principle. The combination of these two values produces the *Likelihood* level which collectively evaluates the probability of a TAI principle being compromised. The *Impact* measures the criticality of the use-case scenario in terms of each TAI principle. It assesses the potential consequences of the principle being compromised within the context of the tool's application.

Figure 2 illustrates the proposed approach is organized sequentially into four steps: *Data Collection*, *Data Modeling & Analyzing*, *Risk Evaluation*, and *Suggestion* which operates in two modes: user-only (M1) or user-plus developer (M2). The figure employs a color-coded system to differentiate between the specific actions and processes associated with each mode: elements highlighted in blue pertain to the *User*, those in green correspond to the *Developer*, and the components in black apply to both modes. Below, we explain each step concisely.

**Data Collection.** The data collection process is going to be performed by having comprehensive questionnaires that cover multiple factors regarding the development of AI tools. Depending on the involvement of the AI developer, three different questionnaires are provided—i.e., *Q1-TAI Implementation*, *Q2-Criticality*, and *Q3-Algorithmic*.

**Data Modeling & Analysis.** The results obtained from the questionnaires in the previous step flow into this step as essential inputs. Based on the scenario mode, out of this step, two models can be generated; (i) the Basic model, which considers *M1* mode, and (ii) the Advanced model, which is enriched with the involvement of both the AI developer and the user. The Advanced model extends beyond user feedback by integrating technical insights, allowing for a more intricate analysis of the AI tool's alignment with TAI principles. There are different automated processes in this step that are connected to each obtained response for the questionnaires, namely, CE Assessment (P1), ABT Assessment (P2), Algorithmic Estimation (P3), and Criticality Analysis (P4). Below, we provide a brief description of each process; **P1**. This process analyses responses to *Q1*, determining CE levels for each TAI principle. For each principle, specific properties are identified (as depicted in Figure 1), with each property being assessed through a series of targeted questions. **P2**. To conduct this analysis, preliminary we need to identify the algorithm used in the AI system. In *M2* mode, this identification is straightforward as the developer specifies the algorithm. In *M1* mode, two scenarios arise: if the tool's documentation is available and the user can specify its algorithm; if not or the user is unable to specify the algorithm, the user is prompted to complete *Q3*, which is part of the subsequent P3 process. **P3**. This process performs in the case of *M1* mode, which helps us uncover the algorithm through responding to *Q3*. The responses obtained from *Q3* determine if the algorithm is rule-based or ML-based. **P4**. For this analysis, the user's responses to *Q2*. We made a correlation between each question in *Q2* and TAI principles (they are constant in our approach), which aids in assessing the extent to which the principles of TAI may be affected in light of the specific use-case scenarios provided by the user.

**Risk Evaluation.** In this step, we conduct likelihood

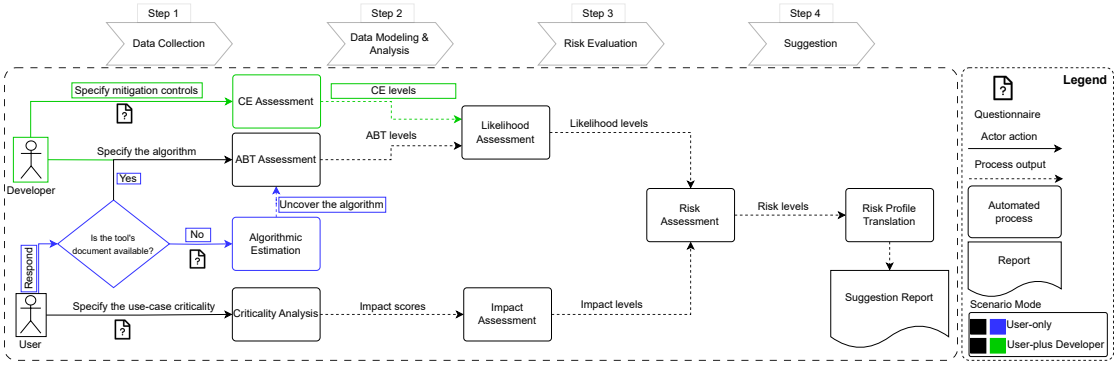


**Figure 2:** The proposed risk-aware approach.

and impact assessments based on the previous step output. Depending on the mode, the risk assessment yields varying risk levels. In fact, the difference between these models lies in the input they provide for assessing likelihood. The basic model operates under constraints due to a lack of developer involvement, overlooking both (i) detailed algorithmic insights, where it might be possible the document of the tool is not available or the user may be unable to extract information regarding the algorithmic information of the tool and only rely on a general estimation, and (ii) CE levels. Instead, the advanced model integrates insights from both actors, providing a comprehensive perspective on the AI tool's trustworthiness.

**Suggestion.** Upon completing the risk assessment step with either the basic or advanced model, the next step is translating the risk profiles into concrete suggestions. This step aims to empower legal practitioners with (actionable) insights to enhance their awareness regarding the trustworthiness and reliability of the AI tool within their judicial workflows.

## 5. Conclusion

We proposed a risk-based approach that offers a systematic method for evaluating and managing potential risks associated with AI applications in judicial contexts. Combining user feedback, particularly from judges, with technical insights, our approach assesses the alignment of AI tools with TAI principles. Through this semi-automated process, we aim to enhance awareness and accountability in AI usage within legal frameworks.

## Acknowledgments

This work was partially supported by the JuLIA project, funded by the Justice Programme of the European Union — JuLIA (101046631), JUST – 2021 JTRA.

## References

[1] B. Li, P. Qi, B. Liu, S. Di, J. Liu, J. Pei, J. Yi, B. Zhou, Trustworthy ai: From principles to practices, ACM Computing Surveys 55 (2023) 1–46.

[2] L. Cobianchi, J. M. Verde, T. J. Loftus, D. Piccolo, F. Dal Mas, P. Mascagni, A. G. Vazquez, L. Ansaloni, G. R. Marseglia, M. Massaro, et al., Artificial intelligence and surgery: ethical dilemmas and open issues, Journal of the American College of Surgeons 235 (2022) 268–275.

[3] A. Reichman, Y. Sagy, S. Balaban, From a panacea to a panopticon: the use and misuse of technology in the regulation of judges, Hastings LJ 71 (2019).

[4] L. Winmill, Technology in the judiciary: One judge's experience, Drake L. Rev. 68 (2020) 831.

[5] W. De Mulder, P. Valcke, J. Baeck, A collaboration between judge and machine to reduce legal uncertainty in disputes concerning ex aequo et bono compensations, Artificial Intelligence and Law 31 (2023) 325–333.

[6] H. AI, High-level expert group on artificial intelligence, 2019.

[7] B. Li, P. Qi, B. Liu, S. Di, J. Liu, J. Pei, J. Yi, B. Zhou, Trustworthy ai: From principles to practices, ACM Computing Surveys 55 (2023) 1–46.

[8] B. Stanton, T. Jensen, et al., Trust and artificial intelligence, preprint (2021).

[9] L. N. Tidjon, F. Khomh, The different faces of ai ethics across the world: a principle-implementation gap analysis, arXiv:2206.03225 (2022).

[10] J. M. Wing, Trustworthy ai, Communications of the ACM 64 (2021) 64–71.

[11] D. Kaur, S. Uslu, K. J. Rittichier, A. Durresi, Trustworthy artificial intelligence: a review, ACM Computing Surveys (CSUR) 55 (2022) 1–38.

[12] S. Thiebes, S. Lins, A. Sunyaev, Trustworthy artificial intelligence, Electronic Markets 31 (2021).

[13] L. N. Tidjon, F. Khomh, Never trust, always verify: a roadmap for trustworthy ai?, arXiv:2206.11981 (2022).

[14] G. Yang, Q. Ye, J. Xia, Unbox the black-box for the medical explainable ai via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond, Information Fusion 77 (2022) 29–52.

[15] T. A. Abdullah, M. S. M. Zahid, W. Ali, A review of interpretable ml in healthcare: taxonomy, applications, challenges, and future directions, Symmetry 13 (2021) 2439.