Cui bono? Judicial decision-making in the era of AI: A qualitative study on the expectations of judges in Germany

Anna-Katharina Dhungel* 1, Moreen Heine1

Abstract • Despite substantial artificial intelligence (AI) research in various domains, limited attention has been given to its impact on the judiciary, and studies directly involving judges are rare. We address this gap by using 20 in-depth interviews to investigate German judges’ perspectives on AI. The exploratory study examines (1) the integration of AI in court proceedings by 2040, (2) the impact of increased use of AI on the role and independence of judges, and (3) whether AI decisions should supersede human judgments if they were superior to them. The findings reveal an expected trend toward further court digitalization and various AI use scenarios. Notably, opinions differ on the influence of AI on judicial independence and the precedence of machine decisions over human judgments. Overall, the judges surveyed hold diverse perspectives without a clear trend emerging, although a tendency toward a positive and less critical evaluation of AI in the judiciary is discernible.

Cui bono? Richterliche Entscheidungsfindung in Zeiten von KI: Eine qualitative Untersuchung zur Erwartungshaltung von Richter*innen in Deutschland


Keywords • artificial intelligence, algorithmic judges, user-centered studies, e-justice, expert interviews

This article is part of the Special topic “AI for decision support: What are possible futures, social impacts, regulatory options, ethical conundrums and agency constellations?,” edited by D. Schneider and K. Weber. https://doi.org/10.14512/tatup.33.1.08

Introduction

While significant strides have been made in the exploration of AI’s impact on various sectors, the judicial domain remains relatively understudied within this discourse. Existing research primarily centers on legal issues related to AI implementation in courts, public perceptions of algorithmic judges, and isolated technical case studies (Eidenmüller and Wagner 2021; Watson et al. 2023; Yalcin et al. 2023). Particularly, the deployment of risk assessment systems in criminal proceedings attracts scholarly attention, notably concerning issues of justice and discrimination, alongside the technical feasibility (Berk 2019; Dressel and Farid 2018; Završnik 2020). However, there is a conspicuous lack of studies that focus on the attitudes of judges towards AI. Publications that specifically address this audience are rare,
both in the context of Germany, and in comparison with other nations. Notably for the German-speaking region, Hartung et al. (2022) examined the future of digital justice, involving interviews with judges, and a publication by IBM compiles insights garnered from discussions with (vice-)presidents of various courts (IBM Deutschland 2022).

Judges provide unique insights into the current state of the art regarding technology use within courts, and they are the target group of the AI systems under consideration. Therefore, their perspectives serve as a critical touchstone for understanding the potential implications arising from AI’s integration into legal proceedings. Building on this premise, through the conduct of 20 in-depth interviews with German judges, this explorative research aims to shed light on the following research questions:

- **RQ1:** According to the surveyed judges, how will a court proceeding appear in the year 2040, and to what extent is AI expected to be involved?
- **RQ2:** What implications would an increased use of AI have on the role of judges and the principle of judicial independence?
- **RQ3:** Based on responses from the judges, should the results and decisions of an AI system prevail over human judgments if the system demonstrably arrives at superior verdicts?

The findings demonstrate a general expectation for the ongoing digitalization of courts, while scenarios for the implementation of AI are only partially conceivable. Concerning the impact of AI on judicial independence, contrasting views were prevalent. Many individuals hold reservations about fully delegating decision-making to machines, perceiving it as both inconceivable and worrisome. Conversely, a portion of respondents advocate for machine-mediated decision-making, contingent upon substantiated evidence demonstrating its superior decision-making capabilities. Overall, the perspectives and views of the surveyed judges are diverse and a clear trend cannot be determined. However, there exists a tendency to evaluate AI implementations in the judiciary more optimistically and positively rather than critically.

**AI in the judiciary: state of the art**

In many countries, it is expected that AI use in legal proceedings will increase in the future. This sentiment is exemplified in China, where an extensive network of AI applications is set to be deployed by 2025, designed to bolster and streamline legal processes (Yu 2022). The United Nations Educational, Scientific and Cultural Organization envisions a rising adoption of AI in the judiciary, evident in the development of a dedicated online course titled “AI and the Rule of Law: Capacity Building for Judicial Systems” (UNESCO 2023). Additionally, a growing demand for the judiciary to go digital has been fueled by citizens’ higher expectations, increased court workloads, succession challenges, and the need to balance the playing field with legal tech providers (IBM Deutschland 2022).

However, gazing into the future is unnecessary, as AI systems are already being formally employed by judges. Risk assessment tools are perhaps best recognized: The objective of these tools is to determine the prospective likelihood of recidivism among offenders. In 49 out of 50 US states, such systems are applied to assess aspects like bail, parole, pretrial custody status, or the duration of sentences (Stevenson 2018). The Chinese AI-driven system ‘Little Judge Bao’ goes further, proposing tailored sentences based on pre-selected factors (Shi 2022).

Looking at the state of digitalization, Singapore serves as a notable instance of a highly digitized judiciary with an all-encompassing online case management system across jurisdictions, facilitating case initiation, monitoring, and data for predictive caseload analysis. Canada showcases another example, launching its first online tribunal in 2012, where all court interactions occur digitally (Hartung et al. 2022).

Germany’s judiciary has fallen behind both internationally and compared to other sectors in adopting digital transformation. According to Hartung et al. (2022), the technological solutions implemented within the German judicial system are limited, outdated, and not sufficiently aligned with user requirements. They estimate that the digitalization of the German judiciary lags behind leading countries by approximately 10–15 years. Dreyer and Schmees (2019) conclude that the feasibility of AI in the judiciary fails solely due to the insufficient availability of training data. Despite this lag in digitalization, the use of algorithms in courts is subject of a critical debate among legal scholars. This encompasses discussions on how algorithms could be deployed within the judiciary to address the shortcomings of human decision-making (Nink 2021), the legal evaluation of so-called ‘robot judges’ (Greco 2020), or the implications of AI deployment on human rights (Završnik 2020). In the field of information systems, the topic has received less attention thus far. Some studies examine the recidivism prediction algorithms already in use in the United States, focusing on aspects such as fairness and reliability (Berk 2019), or the effects of human-machine interaction in this context (Grgić-Hlača et al. 2019). The
target audience of AI systems in the judiciary is predominantly not directly involved in these studies.

Research method

Sample characteristics
The sample (n = 20) was recruited through email invitations to courts, to the Deutscher Richterbund (German association of judges), and through personal networks. It comprises eleven male and nine female individuals, with an average experience of 13.6 years as judges (sd = 10.3 years). Almost all participants hold active judge positions, with only one individual having ceased working as a judge in 2017. The distribution across judicial levels includes 10 judges from local courts, 8 from regional courts, and 2 from higher regional courts. Regarding age, one of the participants is below 30 years old, seven are between 30 to 39 years old, five are between 40 to 49 years old, five are between 50 to 59 years old, and two are above 60 years old. Nine individuals specialize in civil law, three in criminal law, with four of them holding active judge positions in both civil and criminal law. Two participants each practice administrative law and labor law. Participants responded to the Affinity for Technology Interaction (ATI) Scale (Franke et al. 2019). The results, obtained on a scale ranging from 1 (low affinity) to 6 (high affinity), reveal that, as a group, participants demonstrate a moderate level of affinity for technology interaction, with a mean score of \( m = 3.47 \) (sd = .94, range: 2.00–5.00) and a high internal consistency (\( \alpha = .92 \)). This suggests that the sample is not biased by a strong affinity for technology, which could have been possible since participation in the interviews was voluntary, implying an inherent interest in the topic.

Conducting the interviews
This study adopted the reporting framework, guidelines, and dramaturgical model proposed by Myers and Newman (2007) for conducting interviews within the context of information system research. They suggest incorporating essential meta-data related to the interviews (see table 1).

The initial interview was conducted in person, while all subsequent interviews were conducted virtually. The guiding questionnaire consisted of 29 questions, categorized into six sections: current technology usage, AI system requirements, personal attitudes, expectations, human judges’ capacity, and ethics. The present paper emphasizes the questions within the expectations and ethics categories.

Nearly the entire interview process was audio-recorded. In the initial twelve interviews, recording was omitted for the categories human judges’ capacity and ethics, opting for written notes instead. This approach was intended to foster greater trust and enhance participants’ confidence, given the sensitive nature of these questions. However, this method did not produce the desired outcome. As a result, for the subsequent eight interviews, the entire interview process was recorded. The content analysis that followed used the MAXQDA software.

Qualitative analysis
The content analysis was guided by the methodological frameworks put forth by Kuckartz and Rädiker, encompassing both their general approaches and the specific techniques employed for analyzing interviews (Kuckartz and Rädiker 2019; Rädiker and Kuckartz 2020). For the development of the coding scheme, a data-driven approach (inductive methodology) was adopted, involving a step-by-step coding process where codes were iteratively generated until saturation was achieved. The aim is to structure the content and analyze it on the basis of this structure. In this way, diverse attitudes and opinions can be identified. The two authors initially independently coded three randomly selected interviews. The results were then discussed and harmonized. Subsequently, the remaining interviews were independently coded, and their outcomes, such as their alignment with existing codes or the creation of new codes, were deliberated upon. The overall categorical system consists of 72 codes and a total of 339 text passages were coded.

Results

The judiciary in 2040
The judges were asked how a court proceeding might look in the year 2040. The question was received differently, yielding a wide range of responses, as indicated by the numerous codes generated (25 in total). These responses can be categorized into two main themes: future scenarios that describe expectations for forthcoming court proceedings, and critical aspects and concerns regarding the anticipated developments. The following list summarizes the mentioned scenarios, with the frequency of each mention indicated in parentheses.

- Advanced digitalization (17): It was mentioned that “improved technology”, “no more paper files”, and the ability
to “better process documents with computers” are anticipated. Moreover, parties are expected to “communicate electronically with each other”, “submit a relatively significant amount electronically”, and “hopefully, we will indeed have an electronic case file”. Digital documentation is expected to increase, for example, due to the recording of hearings, and it is anticipated that evidence will be technologically processed and digitally accessible.

- **Video hearing (14):** The judges are confident that video conferencing for hearings will continue to expand, anticipating increased opportunities in this regard.

- **Hardly any change (7):** Some judges do not anticipate significant changes within the judiciary, and if any changes are foreseen, they are mostly limited to minor developments in digitalization.

- **AI deployment (5):** Certain judges anticipate an increased use of AI, for instance, in tasks such as suggesting applicable norms for a case, organizing precedent cases, automating the creation of basic legal documents, and employing predictive analysis tools.

- **Pre-trial proceedings (5):** Some judges anticipate the establishment of pre-trial proceedings such as the implementation of predictive tools, resulting in the avoidance of less promising lawsuits, the use of algorithms developed by companies or other private sector stakeholders to decide specific disputes, and the existence of online dispute resolutions.

In addition, the following aspects were addressed, which characterize the scenarios in more detail:

- **Decision-making remains with humans (4):** Four respondents are certain that the ultimate decision will remain with the judge, as one individual expressed, “I believe we still need judges who make the final decision”.

- **Reduced processing times (4):** It is anticipated that, among other factors, the implementation of new technologies will enable court proceedings to be conducted in a shorter span of time.

- **Change in organizational structure (3):** An “infrastructure reform” is anticipated, and it is also assumed that there will be fewer courts and judges as well as fewer case numbers, particularly in civil law.

Furthermore, the following statements were mentioned once each: the potential for new citizen-court interactions, such as online lawsuit filings; the potential partial replacement of judges; the potential use of Virtual Reality; the emergence of digital lawyers for defendants; and the anticipated collaboration enhancement within the EU, possibly facilitated through an EU-wide shared database for decisions.

Regarding the mentioned concerns and critical considerations, it was noted six times that face-to-face conversation is irreplaceable, and four times it was emphasized that human interaction cannot be substituted, with one person saying: “What distinguishes judicial decisions and court proceedings at their core, however, is the personal conversation and the individual context within a legal process. I believe that this cannot be replaced by AI systems because there is a significant amount of social interaction involved, which may not directly relate to legal matters but nonetheless significantly shapes the situation.” Two individuals stated that they believe older judges will struggle with the growing digitalization. Another two highlighted the importance of a societal debate about the use of AI in the legal system, questioning whether we as a society desire such developments. Two respondents expressed concerns about the increasing reliance on technology. The following concerns were raised once each: the growing digital asymmetry within the legal profession, IT security, the lack of competence of IT service providers, and concerns about the rule of law.

### The future role of judges

Subsequently, participants were asked about their expectations regarding the development of the judge’s role by 2040. The responses varied between positive expectations, concerns, and neutral statements (see table 2).

<table>
<thead>
<tr>
<th>Optimistic Anticipation</th>
<th>£</th>
<th>Concerns</th>
<th>£</th>
<th>Neutral</th>
<th>£</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relief through digitalization</td>
<td>7</td>
<td>Reduced decision-making authority</td>
<td>2</td>
<td>No changes of the judge’s role</td>
<td>6</td>
</tr>
<tr>
<td>AI as support and assistance</td>
<td>4</td>
<td>Reduced reverence</td>
<td>1</td>
<td>Judges as case managers and mediators</td>
<td>5</td>
</tr>
<tr>
<td>AI in mass proceedings</td>
<td>1</td>
<td>Rise in information overflow</td>
<td>1</td>
<td>Judgment remains with the human</td>
<td>3</td>
</tr>
<tr>
<td>Additional responsibilities</td>
<td>1</td>
<td>New competencies necessary</td>
<td>3</td>
<td>Surveillance of the systems</td>
<td>2</td>
</tr>
</tbody>
</table>

**Tab.2:** Categorization of potential changes of the judge’s role by 2040.

*Source: interviews conducted by the authors*
The question was raised as to whether the implementation of AI systems within the judiciary might lead judges to excessively rely on them, potentially fostering automation bias – a tendency to overly trust automated systems, which may result in errors or overlooking something (Skitka et al. 2000). This question specifically pertained to decision support systems, where the human makes the final decision, for instance, pre-filing court orders. According to Sheridan’s automation scale – ranging from 1 = the human must decide and execute everything to 10 = the system acts autonomously and decides without human involvement – these systems fall within levels up to a maximum of 5 (Sheridan et al. 1978).

Some judges noted this issue persists even without AI, such as when they agree to a prosecutor’s case dismissal request to reduce workload (mentioned in 4 interviews). Younger individuals, with greater trust in technology, were identified by two respondents as more prone to agree with the system, reinforcing automation bias. Additionally, it was noted that judges often face time constraints, which could lead them to go along with the system’s decision simply due to time pressure (4). Proposed solutions included the need for judges to receive appropriate training (1), designing the systems and their usage context with psychological incentives to avoid automation bias (1), and the implementation of relevant regulations (1). In contrast, some respondents (5) believe that automation bias is not a problem for judges because they are “self-disciplined”, that they have “inherent skepticism towards anything that touches their own high decision-making authority”, and, ultimately, that the “professional group is inherently inclined to resist”.

Participants were also asked whether judicial independence is called into question with an increased use of AI. On the one hand, some (5) believe the development to be critical due to concerns about a gradual takeover by such systems, with judges increasingly facing pressure to justify decisions that do not align with those of AI, and fears that AI might render judges redundant. On the other hand, the argument made was that judicial independence is not at risk, as ultimately, judges decide how and when to use technology, with AI systems serving merely as assistants (7). Moreover, it was emphasized that a threat to judicial independence would depend on whether the use of AI systems would be mandatory and how the integration of such algorithms into procedural rules would occur (12). Additionally, it was stressed twice that it is within the responsibility of individual judges to determine whether their own independence would be compromised or not: “I believe that an AI system can pose a significant threat to lazy-minded judges.”

### Perspectives on AI-generated judgments

During the interviews, the judges were also asked: Should AI system outcomes override human judgments when the system consistently yields better verdicts? In this context, the discussion pertained to systems classified at level 10 on the automation scale, meaning they operate autonomously without human intervention (Sheridan et al. 1978). The majority of respondents (14) initially countered by stating that it is not demonstrable at what point a decision would be considered “better”. Subsequently, responses to this question diverged significantly. Some supported the idea of machines issuing judgments, while others endorsed it only under certain conditions or for specific use cases. Conversely, many responses entailed explicit and absolute rejection (see table 3).

As evident from the frequency of mentions, it is apparent that not only were the respondents divided in their opinions, but individual participants also provided varying statements. Three statements not included in the table indicate that the responsibility of AI systems for decision-making is a societal choice: “I believe that, since we live in a democracy, if society decides that we want this, it should be done.”

### Discussion

Judges operate with autonomy, determining their operational methodologies, and hold accountability for each procedural facet, as articulated in Art. 97 Abs. 1 GG, which underscores their independence and subordination solely to the law. They typically lack dedicated secretarial support or personally assigned assistants, exemplifying the self-directed nature of their role. Consequently, this engenders, on the one hand, the fundamental latitude for judges to exercise discretion in adopting technologi-
cal aids, unless statutory provisions dictate otherwise. On the other hand, it underscores the challenge of seamlessly integrating AI-based assistance systems into existing judicial processes.

It is to be noted that the interviews do not reveal a consistent consensus; a multitude of diverse viewpoints were expressed. However, it is noticeable that there is a general tendency towards a more favorable outlook rather than a critical one regarding the future possibilities of AI application, such as its potential use as a helpful tool in mass proceedings or as an assistance system for case processing. At the same time, critical topics such as IT security or data for such systems and the associated potential for discrimination were scarcely addressed. The slightly positive view held by judges could be attributed to their potential lack of AI expertise compared to members of, for instance, the information systems community, leading to a limited understanding of the technological challenges associated with AI. Judges tend to perceive digitalization and related AI technologies as advantageous for their daily work, hence the positive outlook.

Regarding the first research question concerning expectations for the year 2040, the responses demonstrate a strong reliance on digitalization. This underscores the previously mentioned lag in the German judiciary and the judges’ expectations that this gap will be bridged in the coming years. Increased AI deployment is only expected to a limited extent. Regarding the second research question concerning potential impacts on the role of judges, on the one hand, a positive expectation was revealed, such as relief through digitalization. On the other hand, concerns were expressed, for instance, regarding reduced decision-making authority.

A similar pattern emerged in response to the question about judicial independence, with some expressing concerns about a gradual takeover by AI, while others had no reservations. The risk of automation bias coming with regular use of such systems, in turn, was largely acknowledged. Concerning the third research question about whether judgments from AI systems should potentially be deemed more significant than those made by humans, there were supporters who could envision such a scenario under specific circumstances, as well as opponents who assert that such an outcome is precluded. Notably, the diversity of responses to the previous questions remained evident in addressing this question as well, although there is research that demonstrate that higher levels of automation are frequently met with less acceptance when compared to lower levels (Ghazizadaeh et al. 2012).

The study has notable limitations to consider. It is confined to the German context, potentially impacting its applicability to other legal systems. Due to the small sample size, the study is not representative. Also, the judges’ self-selection might introduce bias, as they hold an interest in AI. Finally, different interpretations and confusion regarding AI and digitalization were observed. Despite the predominantly descriptive nature of the analysis, it might serve as a valuable resource for future research endeavors, particularly for theory building.

Outlook

According to the draft of the EU AI Act (Article 8 of Annex III), AI systems used in the judiciary are classified as high risk. Therefore, the deployment of AI systems for judges is already politically viewed with skepticism. As a result, scenarios involving the use of AI give rise to various legal, technical, and ethical questions, such as: How can the outcomes of AI systems be made comprehensible for judges (encompassing the broad extensive topic of explainable AI)? How can procedural justice and the right to a fair hearing be ensured? How can ongoing legal oversight be maintained despite the use of AI, and self-reinforcing processes be prevented? What specific impact do particular systems have on the decision-making of judges?

It is an ongoing societal debate; therefore, it is essential that scientific research is conducted to ensure the effective customization of solutions to the distinct requirements of judges. The partnership between legal scholars and computer scientists becomes pivotal in cultivating approaches that address the unique demands of a contemporary judicial system. Future interdisciplinary research should focus on exploring in a human-centered manner how judges can effectively employ AI in ways that align with technical feasibility, streamline their work processes, and gain societal acceptance.

Acknowledgement

We thank Eva Beute for her support during the interviews and we thank the reviewers for their constructive and supportive feedback.

Funding • This work received no external funding.

Competing interests • The authors declare no competing interests.

References


Nearly one-third of the respondents anticipate that the role of the judge will not change.


Kuckartz, Udo; Rädiker, Stefan (2019): Analyzing qualitative data with MAXQDA. Text, audio, and video. Cham: Springer.


ANNA-KATHARINA DHUNGEL holds a Master’s degree in International Politics and International Law, as well as a Master in Business Informatics. With five years of experience as an IT Consultant, she joined the University of Lübeck in 2021 to pursue her PhD and conduct research on the implementation of artificial intelligence in the field of justice.

PROF. DR. MOREEN HEINE has been a professor of E-Government and Open Data Ecosystems at the University of Lübeck and scientific director of the Joint Innovation Lab since 2019. Previously, she was an assistant professor of business information systems at the University of Potsdam. She is a National E-Government Competence Centre (NEGZ) board member.