UNSW Law & Justice Research Series

Generative Artificial Intelligence and Article 6 of the European Convention on Human Rights: The Right to a Human Judge?

Mimi Zou

UNSW Law & Justice
UNSW Sydney NSW 2052 Australia

Generative Artificial Intelligence and Article 6 of the European Convention on Human Rights: The Right to a *Human* Judge?

Mimi Zou and Ellen Lefley*

In Mimi Zou, Martin Ebers, Cristina Poncibò and Ryan Calo (eds), *The Cambridge Handbook of Generative AI and the Law* (Cambridge University Press 2025)

Keywords: generative artificial intelligence; large language models; judicial use of artificial intelligence; right to a fair trial; European Convention on Human Rights; human dignity; access to justice.

1   Introduction

The public launch of ChatGPT in December 2022 sparked widespread excitement and concern about AI's potential and risks. The powerful capabilities of the latest generative AI models have captured public attention. For every article or social media post praising AI as a revolutionary solution to major challenges, there is another warning of its existential threat to humanity. Policymakers across various sectors are grappling with the same question: how can we harness the transformative benefits of AI while mitigating its risks?

The justice sector is no exception to the impact of AI. Regardless of whether one views AI as a transformative opportunity or an existential threat, it is essential to anticipate how new and emerging AI technologies are changing the justice system. AI is already being used in courtrooms worldwide, with lawyers and self-represented litigants using chatbots and other AI tools for research and drafting submissions. Judges, judicial officers, and support staff are also starting to utilise AI in their work, with judicial guidelines issued in several jurisdictions since 2023.[1]

For some, there has been a strong desire for the adoption of AI in the justice sector due to longstanding concerns about the inaccessibility of justice and the significant inequality between those who can afford lawyers and those who cannot. Consequently, even though not all risks have been identified, tested, or assessed, there is a compelling attraction to using AI chatbots for legal advice and drafting, as they are arguably far better than having no assistance at all.

[1] See Courts and Tribunals Judiciary, England and Wales, 'Artificial Intelligence - Judicial Guidance' (12 December 2023) www.judiciary.uk/guidance-and-resources/artificial-intelligence-ai-judicial-guidance/; Courts of New Zealand, 'Guidelines for use of generative artificial intelligence in Courts and Tribunals' (2023) www.courtsofnz.govt.nz/going-to-court/practice-directions/practice-guidelines/all-benches/guidelines-for-use-of-generative-artificial-intelligence-in-courts-and-tribunals/; Hong Kong Judiciary, 'Use of Generative AI for Judges and Judicial Officers and Support Staff of the Judiciary' (2024) www.judiciary.hk/doc/en/court_services_facilities/guidelines_on_the_use_of_generative_ai.pdf.; Supreme Court of New South Wales, 'Guidelines for New South Wales Judges in Respect of Use of Generative AI' (2024) https://supremecourt.nsw.gov.au/documents/About-the-Court/policies/Guidelines_Gen_AI.pdf.

Access to justice issues extend beyond just the affordability of legal advice. With record levels of court backlogs,[2] it has been a common occurrence for people to endure years-long delays in their cases. Looking to the immediate future, as more people are assisted by AI tools to understand, articulate, and assert their legal rights, the irony is that increasing access to advice may exacerbate the problem of delay. Indeed, the already overrun resources of the justice system may be stretched further, creating even greater strain. Therefore, the complexity of these challenges focuses our attention not only on how generative AI can help the public to help identify and express their legal problems, but also how it can assist in resolving them. This includes broader consideration of generative AI's role in dispute prevention, as well as alternative dispute resolution (ADR) and court proceedings.

Importantly, the potential for people to have legal problems resolved primarily or entirely by AI systems in the judicial setting is one that stirs up considerable debate. The Council of Europe's Ethical Charter on the use of AI in judicial systems identified one of the core principles to be respected as the need to ensure 'guarantees to the right of access to the judge'.[3] The Council of Bars and Law Societies of Europe also published in May 2023 a policy statement on AI and the justice system, stating 'a right to a human judge should be guaranteed at all stages of the proceedings'.[4]

On 13 March 2024, the European Parliament adopted the Artificial Intelligence Act (AI Act).[5] The AI Act came into effect in August 2024, marking the first legal regulation of AI in the EU. Annex III, Article 8(a) of the AI Act classifies certain judicial use of AI as 'high-risk':[6]

> 8. Administration of justice and democratic processes:
> (a) AI systems intended to be used by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts, or to be used in a similar way in alternative dispute resolution;

This is elaborated in Recital 61:

> Certain AI systems intended for the administration of justice and democratic processes should be classified as high-risk, considering their potentially significant impact on democracy, the rule of law, individual freedoms as well as the right to an effective remedy and to a fair trial. In particular, to address the risks of potential biases, errors and opacity, it is appropriate to qualify as high-risk AI systems intended to be used by a judicial authority or on its behalf to assist judicial authorities in researching and interpreting facts and the law and in applying the law to a concrete set of facts… The use of AI tools can

---

[2] UK Government statistics show delays have increased since the pandemic rather than recovered. For example, the average time for civil small claims to reach trial in England and Wales was 39 weeks in September 2019, which rose quickly to above 50 weeks during the pandemic. This figure has since reached 55.4 weeks in September 2023. See His Majesty's Courts and Tribunal Services, 'HMCTS Management Information' (National data on workloads and timeliness, His Majesty's Government, 2024) www.gov.uk/government/collections/hmcts-management-information. For criminal justice delays, see also National Audit Office, 'Reducing the backlog in the Crown Court Report – Value for money' (UK report on the Ministry of Justice court backlogs, 24 May 2024) www.nao.org.uk/reports/reducing-the-backlog-in-the-crown-court/.

[3] Principle One of the Council of Europe, 'European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment' (Adopted at the 31st plenary meeting of the European Commission for the Efficiency of Justice in Strasbourg, 3-4 December 2018) at 8.

[4] Council of Bars and Law Societies of Europe, 'CCBE Statement on the use of AI in the justice system and law enforcement' (25 May 2023) at 1 www.ccbe.eu/fileadmin/speciality_distribution/public/documents/Statements/2023/EN_ITL_20230525_CCBE-Statement-on-the-use-of-AI-in-the-justice-system-and-law-enforcement.pdf.

[5] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence, OJ L 2024/1689.

[6] Annex III lists high-risk AI systems, pursuant to Article 6(2) of Regulation (EU) 2024/1689.

support the decision-making power of judges or judicial independence, but should not replace it: the final decision-making must remain a human-driven activity.

In this context, a crucial question arises as to how judicial use of AI, especially considering the rapid advancements in generative AI, affects the right to a fair trial. This chapter examines the right to a fair trial as outlined in Article 6 of the European Convention on Human Rights (ECHR),[7] an international human rights treaty that covers 46 countries in Europe and has given rise to an extensive body of case law. Various issues related to the interpretation and application of Article 6 of the ECHR emerge from this examination. Notwithstanding the AI Act's declaration that 'the final decision-making must remain a human-driven activity', our main concern in this chapter is whether Article 6 actually recognises the right to a human judge, with a focus on the civil limb of Article 6.

In addressing this concern, our analysis points to the lack of attention to date to the principle of human dignity in interpreting Article 6 of the ECHR in the context of judicial use of AI. We argue that human dignity necessitates a deeper understanding of Article 6, recognising the intrinsic value of human judgement, empathy, and the importance of being heard and respected by the decision maker. Given courts are public institutions and the ultimate authority in resolving civil disputes, we find that a dignity-based interpretation of Article 6 includes the right to a human judge.

Our chapter is structured as follows. Part 2 analyses the potential implications of using generative AI in the civil justice system for key aspects of Article 6 of the ECHR. These include the right and access to a court within a reasonable time, the impartiality and independence of the judiciary, procedural fairness, and the right to a reasoned judgment. Although existing case law does not explicitly recognise a right to a human judge, Part 3 explores a richer understanding of Article 6 through the lens of human dignity. We conclude that this perspective gives rise to an interpretation of Article 6 that supports the right to a human judge. We also suggest how AI systems should be deployed in civil justice to complement, rather than undermine, human dignity.

While there are various definitions of 'generative AI', our working definition refers to AI systems that generate content, typically in the form of text, images, or other data, based on patterns learned from large datasets. These systems include different models—perhaps the most well-known at the time of writing are large language models (LLMs) on which applications like ChatGPT are based. Yet, we do not consider it helpful to limit our analysis to the generative AI models available at the time of writing. Ever increasing claims are being made about the latest models' capabilities and the reliability of their outputs, sometimes contradictorily even within the same company.[8] We therefore consider a prudent approach that takes into account the following:

i) We cannot predict the scientific breakthroughs that will support new LLMs and other AI models in the near future, let alone over the next 20 years.

---

[7] European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, ETS 5, 4 November 1950.

[8] At the launch of Meta's new LLM, Llama 3, in April 2024, Meta's head of global affairs, Sir Nick Clegg, said it had 'vastly improved capabilities like reasoning'. However, Meta's Chief Scientist Yann LeCunn has made clear his alternative modelling research – so called 'world-modelling' – is progressing precisely because an LLM is not reasoning: 'it certainly appears to most people as reasoning – but mostly it's exploiting accumulated knowledge from lots of training data'. See Hannah Murphy and Cristina Criddle, 'Meta AI chief says large language models will not reach human intelligence', The Financial Times, 22 May 2024 www.ft.com/content/23fab126-f1d3-4add-a457-207a25730ad9.

ii) Even if new models are released that claim to 'understand' or 'reason', we cannot equate these models with human understanding and reasoning.[9] It is important to consider the motivations of companies that may exaggerate their models' capabilities.[10]

iii) Advancements in the development and deployment of this technology in courts will continue to appeal to those seeking solutions for an under-resourced and backlogged justice system.

If an AI-powered chatbot were used to determine a list of small claims today, it would be unsurprising to find some errors in its outputs. These erroneous outputs would be relatively straightforward to analyse: their unreasonable and arbitrary nature is very likely to breach an individual's right to a reasoned judgment under Article 6 of the ECHR. However, a more challenging scenario is when generative AI systems produce outputs in judicial processes that fall within a range of legally plausible outcomes—our focus here is on the impact of such a scenario on Article 6.

2   Article 6 of ECHR and Implications of Generative AI

2.1   Why Article 6 of the ECHR?

As a preliminary point, it is important to explain why Article 6 of the ECHR, and human rights law more generally, deserves our focus in the context of analysing the impact of generative AI on civil justice. Ethical guidelines of the use of AI issued by numerous international organisations have consistently referred to the importance of the protection of fundamental rights generally[11] and the right to a fair trial specifically—for example, one of the key principles of the European Ethical Charter on the use of AI in Judicial Systems and Their Environment refers to the need to ensure that the use of AI must not 'undermine... the right to a fair trial'.[12]

In Europe, enforceable legal standards for automated data processing exist within data protection law. Article 9 of the Council of Europe's Convention 108+ secures the right for an individual 'not to be subject to a decision significantly affecting him or her based solely on an automated processing of [their personal] data without having his or her views taken into Consideration'.[13] This was then developed by the EU, first in their now invalid 1995 Data Protection Directive[14] and later in Article 22 of the General Data Protection Regulation (GDPR).[15] Article 22 of the GDPR gives data subjects 'the right not to be subject to a decision

---

[9] For example, in September 2024, OpenAI released beta o1 advanced reasoning large language models which are 'designed to reason about hard problems using broad general knowledge about the world'. See OpenAI Platform, 'Reasoning Models' (Website, 2024) https://platform.openai.com/docs/guides/reasoning/reasoning.

[10] A Microsoft report in March 2023 titled 'Sparks of Artificial General Intelligence: Early experiments with GPT-4' sparked criticism for creating such hype. An AI scientist unaffiliated with the study was quoted calling Microsoft's report an example of 'big companies co-opting the research paper format into PR pitches'. See Darren Orf, 'AI Has Evolved to Reason Like Humans, Scientists Say', Popular Mechanics, 16 May 2023.

[11] See, for example, Council of Europe, 'European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment' at 7; United Nations Educational, Scientific and Cultural Organisation, 'Recommendations on the Ethics of Artificial Intelligence' at 28.

[12] Principle One of the Council of Europe, 'European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment' at 8.

[13] Article 9(1)(a) of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS No. 108).

[14] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ L 281.

[15] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, OJ L 119.

based solely on automated processing [of their personal data], including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her'.

Data protection laws, however, have their limitations for addressing potential risks arising from the judicial use of AI. Firstly, data protection laws are only engaged when a natural person's personal data is being processed. One can think of many types of civil disputes that involve non-personal information, such as a factual dispute where the data processed concerns the accounts of small businesses or a legal dispute involving the interpretation of case law. Secondly, data protection provisions only prohibit fully automated decision making, and do not address the use of *assistive* generative AI (and provide no protection thereof). Thirdly, there are broad exceptions in data protection law that require a rights framework to navigate. Both Convention 108+ and the GDPR permit exceptions to the prohibition on automated decision making, as long as they safeguard the data subject's 'rights, freedoms, and legitimate interests' and 'respect the essence of fundamental rights and freedoms and constitute a necessary and proportionate measure in a democratic society'.[16] Therefore, in our analysis of the judicial use of generative AI, data protection law does not provide a complete answer.

Elsewhere in EU law, the impact on the right to a fair trial is explicitly referred to in the AI Act's reference in Annex III, Article 8(a) in the context of the use of AI in judicial decision making as 'high-risk'. As discussed earlier, AI systems used by or on behalf of judicial authorities to assist in researching, interpreting, and applying the law to specific facts are classified as 'high-risk'. This classification is driven by concerns regarding the risks of biases, errors, and lack of transparency involved in such deployment of AI systems.[17]

However, Article 6(3) of the AI Act provides exceptions to the high-risk classification, namely where a high-risk AI system referred to in Annex III 'does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making'. According to the second subparagraph of Article 6(3), this exception applies to situations in which:

> (a) the AI system is intended to perform a narrow procedural task;
> (b) the AI system is intended to improve the result of a previously completed human activity;
> (c) the AI system is intended to detect decision making patterns or deviations from prior decision making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review; or
> (d) the AI system is intended to perform a preparatory task to an assessment.

The AI Act does not define further the circumstances in which these relatively broad exceptions could apply. Potentially significant risks can still flow from such 'exceptions', when generative AI is applied in the judicial function to perform narrow procedural tasks (e.g. automated draft generation for routine court orders), improve previously completed human activity (e.g. AI enhancement of judicial opinions), detect decision making patterns (e.g. generative analysis of judicial decision trends), and perform preparatory tasks (e.g. AI-generated case summaries and research memos). The application of 'exceptions' to high-risk AI is complicated by the technology's capabilities, which are advancing quickly and continuously. What constitutes a 'narrow' task or 'preparatory' work when an AI can generate complex, nuanced text that mimics human-created content? Moreover, the cognitive impact of interacting with AI-generated content in judicial work is not yet fully understood.

---

[16] See Article 9(2) and 11(1) of Convention 108+. Article 22 (2)(b) and (3) of the GDPR indicates that data controllers themselves must safeguard rights, freedoms, and legitimate interests when they are making automated decisions under a contract or subject to consent. If it is pursuant to a member state law to which the controller is subject, the law itself must lay down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests.

[17] Recital 61 of the Regulation (EU) 2024/1689.

As generative AI capabilities rapidly advance, there is a real risk that systems initially classified as exceptions under the AI Act could evolve to have more significant impacts, necessitating ongoing scrutiny. Overall, the provisions of the AI Act, despite their emphasis on human-driven decision making, leave room for interpretation and potential loopholes. The AI Act may therefore be insufficient to ensure appropriate safeguards and oversight of high-risk AI systems used in judicial settings. This is where existing human rights frameworks provide important protections in addressing such gaps. Indeed, Preamble 6 of the AI Act requires that the use and regulation of AI be consistent with other fundamental rights and freedoms.

Therefore, in assessing the impact of judicial AI, we pay specific attention to Article 6(1) of the ECHR. Article 6(1) concerns the right to a fair trial in line with the rule of law in democratic societies. The text of Article 6(1) ECHR reads:

> In the determination of his civil rights and obligations or of any criminal charge against him, everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law. Judgement shall be pronounced publicly but the press and public may be excluded from all or part of the trial in the interests of morals, public order or national security in a democratic society, where the interests of juveniles or the protection of the private life of the parties so require, or to the extent strictly necessary in the opinion of the court in special circumstances where publicity would prejudice the interests of justice.

Uses of generative AI in judicial functions and court processes raise a range of issues relating to four important aspects of Article 6(1) of the ECHR, which we examine in this Part: (i) right and access to a court, (ii) independence and impartiality, (iii) procedural fairness, and (iv) right to a reasoned judgment. We note that the actual and potential impacts of generative AI deployment are likely to vary across different types of civil cases. As such, while we examine these aspects of Article 6(1) in light of potential opportunities and concerns, our analysis is by no means exhaustive, especially given the ongoing development and deployment of generative AI tools.

2.2   Right of access to a court

Perhaps the most appealing argument for those advocating for greater use of AI in the civil justice system concerns a central tenet of Article 6(1), which is access to a court. This mandates the provision of an effective judicial recourse for litigants to assert their civil rights.[18] The right of access to a court must be 'practical and effective', and not 'theoretical and illusory'.[19] In specific circumstances, the practical and effective nature of this right may be impaired, for instance, by the prohibitive cost of the proceedings in view of the individual's financial capacity, or by issues relating to time-limits including strict deadlines or prolonged appeal processes.[20]

Importantly, the right to access a court includes having one's case heard 'within a reasonable time'. This explicit requirement under Article 6 is based on the 'importance of administering justice without delays which might jeopardise its effectiveness and credibility'.[21] The European Court of Human Rights has pointed out that excessive delays in the administration of justice have affected public confidence in the judicial system and seriously jeopardised the rule of law on which the ECHR was based.[22] Given the significant backlogs of

---

[18] *DH and Others v the Czech Republic* [2007] ECHR 57325/00 (13 November 2007) § 49.

[19] *Bellet v. France* [1995] ECHR 23805/94 (4 December 1995) § 38.

[20] *Melnyk v. Ukraine* [2006] ECHR 23436/03 (28 March 2006) § 26.

[21] *H. v. France* [1989] ECHR 5856/72 (24 October 1989) § 58; *Vernillo v France* [1991] ECHR 11889/85 (20 February 1991) § 38

[22] *Vanden Kerkhof v. Belgium* [2023] ECHR 13630/19 (5 September 2023).

cases that many courts face,[23] it is unsurprising that the European Ethical Charter on the use of AI in Judicial Systems and Their Environment has encouraged the application of AI to improve judicial efficiency and quality.[24] It is here that generative AI can make a difference in expediting the resolution of civil claims, reducing the costs of accessing a court, and alleviating the main barriers that may inhibit individuals from accessing the courts within a reasonable timeframe.

Generative AI-powered chatbots and virtual assistants can provide tailored information, resources and guidance to court users, and even generate and draft relevant legal documents. Such LLM-based tools can offer invaluable support in navigating court processes and initiating or responding to legal proceedings promptly, particularly for self-represented litigants. The transcription capabilities of generative AI models potentially enable high-quality and efficient record-keeping of court proceedings, ensuring real-time documentation and facilitating post-hearing analysis. Furthermore, improved language accessibility for court users arises from LLMs' powerful translation capabilities and the ability to simplify complex legal language. By streamlining various court processes, generative AI can potentially reduce the overall cost and duration of legal proceedings, enabling many individuals to access a court within a reasonable timeframe.

The above applications of generative AI in civil justice processes are unlikely to be controversial, though they are not without concerns. In addition to data security and privacy issues, it is crucial to ensure fair and equitable access for all court users. Disparities in access could worsen existing inequalities in the justice system, undermining trust in the judiciary and confidence in the legal process, which is already difficult to navigate for many people worldwide.[25]

The more contentious landscape concerns the integration of generative AI in judicial decision making, tapping into its immense potential in expediting fact-finding, legal analysis, and research, potentially alleviating significant case backlogs and reducing onerous judicial workloads.[26] The use of AI in this context necessitates careful consideration of its impact on various aspects of the right to a fair trial, such as those as examined below. A primary concern is the potential for AI-generated errors, such as hallucinations and bias built into AI systems, to cause delay and prolong court proceedings. These risks pose a potential obstacle to the right of access to a court, as they may introduce uncertainties and inaccuracies into the judicial process, risking the erosion of public trust in the legal system. Recognising the risks arising from AI-generated errors, judicial guidelines for the use of AI in court in the UK have specifically recommended that AI should not be used by judicial officers for legal research that cannot be independently verified.[27]

---

[23] This ongoing problem for courts has led to the European Commission for the Efficiency of Justice to develop and adopt a new tool to reduce the backlogs of court cases. See further: CEPEJ, 'The CEPEJ adopts its «Backlog reduction tool»: a new roadmap for identification and addressing structural delays in court proceedings' (16 June 2023) https://www.coe.int/en/web/cepej/-/the-«backlog-reduction-tool»-the-cepej-s-new-roadmap-to-national-judicial-systems-for-identification-and-addressing-structural-delays-in-court-proceedings

[24] Council of Europe, 'European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment' at 5.

[25] JUSTICE, Preventing Digital Exclusion from Online Justice (April 2018) https://files.justice.org.uk/wp-content/uploads/2018/06/06170424/Preventing-Digital-Exclusion-from-Online-Justice.pdf; Victor D. Quintanilla, Kurt Hugenberg, Margaret Hagan, Amy Gonzales, Ryan Hutchings, Nedim Yel, 'Digital Inequalities and Access to Justice' in David Freeman Engstrom (ed) *Legal Tech and the Future of civil Justice* (Cambridge University Press 2023) 225-250.

[26] Sir Geoffrey Vos, 'Speech by the Master of the Rolls to the Bar Council of England and Wales' (20th Annual Law Reform Lecture at Lincolns Inn on Thursday 21 June 2023, Courts and Tribunals Judiciary, 18 July 2023) www.judiciary.uk/speech-by-the-master-of-the-rolls-to-the-bar-council-of-england-and-wales/.

[27] Courts and Tribunals Judiciary, 'Artificial Intelligence (AI) - Judicial Guidance' at 6.

2.3   Independence and Impartiality

Independence and impartiality, as enshrined in Article 6 of the ECHR, are fundamental pillars of the judiciary. The principles are intricately connected and underpin essential democratic and rule of law principles.[28]

First, the independence of a 'tribunal' (terminology in Article 6 that covers a range of court-like institutions) from the executive and the parties is assessed by reference to the manner of appointment of the tribunal's members, the duration of the term of office of its members, the existence of guarantees against outside pressures, and whether the tribunal presents an appearance of independence.[29]

Generative AI has the potential to meet the demands of independence by being designed and operated externally to the executive or litigating parties. At present, much of the technological developments in AI occurs in the private sector. While this private-sector development could potentially insulate AI systems from direct governmental influence, it also introduces new concerns. The risk of indirect influence through funding, development priorities, or data selection in AI training cannot be overlooked. Such influences, even if unintended, could undermine both the actual independence of AI-driven or AI-assisted judicial processes and the public perception of that independence.

To address these concerns and maintain the integrity of judicial independence, robust safeguards and transparency measures are crucial. One approach is to mandate transparent development processes for any AI systems intended for use in courts. This transparency could include full disclosure of funding sources, development methodologies, and design specifications, including open access to the training data. This would serve a dual purpose: it would allow for public scrutiny of potential influences on the AI system, and it would provide a new framework for assessing the 'appointment' and 'tenure' of AI in judicial roles – concepts that traditionally apply to human judges but need reinterpretation for AI systems.

However, implementing such measures is not without challenges. The private companies developing these AI systems may resist, citing concerns about trade secrets and competitive advantages. This resistance highlights the tension between the need for transparency in judicial processes and the proprietary nature of cutting-edge generative AI technology. Albeit not generative AI, the use of the COMPAS algorithm in the US has been a prominent example. Developed and sold by a private company (originally known as Northpointe, now Equivant), the COMPAS algorithm determined a risk score for defendants to be used by judges in the context of criminal sentencing and bail decision making. Its methodology was a trade secret. Its proprietary nature meant that judges could not evaluate how the risk scores were determined or how different factors were weighed by the algorithm to produce the risk score. This led to concerns about the judicial competence required for interpreting the outputs of such systems, and the independence and impartiality of the resulting judicial risk assessment exercise when incorporating such technology.[30] Such concerns were particular salient in light of evidence that Black defendants were more likely to be falsely identified as high risk, while white defendants are more likely to be falsely identified as low risk by the algorithm.[31]

Turning to the requirement of impartiality under Article 6, a tribunal must be free from prejudice and bias. This is assessed in two steps. First, a subjective test necessitates that the tribunal be free from bias or prejudice in each case regarding the conviction of behaviour of a

---

[28] *Kleyn and Others v. the Netherlands* [2003] 39343/98, 39651/98, 43147/98, 46664/99 (6 May 2003) § 192.
[29] *Langborger v. Sweden* [1989] 11179/84 (22 June 1989) § 32; *Kleyn and Others v. the Netherlands* § 190.
[30] 'State v. Loomis' (2017) 130 Harv. L. Rev. 1530
[31] Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, 'Machine Bias' and 'How we analysed the COMPAS recidivism algorithm', *Propublica* (23 May 2016).

particular judge. Second, an objective test requires that the tribunal takes sufficient steps to exclude legitimate doubt of its impartiality.[32] Together, then, the requirement of independence and impartiality puts demands on tribunals to not only be free from prejudice, bias and to be independent from the executive and the parties, but also to take steps to avoid the appearance of not having impartiality or independence.

The use of AI in judicial decision-making is sometimes represented as an opportunity to minimise human bias and prejudice, giving rise to perceptions of impartiality. However, this purported objectivity is over-simplified. AI systems are subject to human bias due to numerous points of human contact during their design, development, maintenance, and use.[33] As just one example during use, phrasing of questions or prompts can skew the outputs generated.[34] Bias and prejudice may also emerge from the training data and data sources available for the training of LLMs. This includes the use of training data that features historical bias and/or inequalities or is missing important data points.[35] As discussed further in Part 3, the risk of AI bias may not only undermine judicial impartiality but may also give rise to the potential of more exaggerated forms of discrimination, objectification, and dehumanisation of individuals before the court.

2.4   Procedural fairness

Procedural fairness is a multi-faceted requirement of Article 6(1). As a starting point, the principle of legal certainty constitutes one of the basic elements of the rule of law underpinning procedural fairness.[36] The use of generative AI in judicial decision making may help to improve consistency in outcomes in similar cases, drawing on vast data related to judicial trends.

Another key component of procedural fairness encompasses the right of parties to present observations which they regard as relevant to their case. Generative AI tools can support the preparation and presentation of submissions, arguments and evidence brought forward by parties. Generative AI has capacities to produce and summarise case law, explanations, and articles to support the communication of legal knowledge in accessible forms to litigants. This can clarify an understanding of what might be relevant to bring forward as a litigant in a civil matter. These uses of generative AI may be of particular benefit in the context of fairness for those litigants who may be illiterate or do not speak English.

Procedural fairness also involves the appropriate consideration of the observations by the court and hence, requires courts to conduct proper examinations of submissions, arguments and evidence adduced by the parties.[37] The use of generative AI tools by judicial decision makers in the consideration of submissions, arguments, and evidence adduced by parties raises

---

[32] *Wettstein v. Switzerland* [2000] 33958/96 (21 December 2001) § 43; *Micallef v. Malta* [2009] 17056/06 (15 October 2009) § 93.

[33] Aziz Z. Huq, 'A Right to a Human Decision' (2020) 10(3) *Virginia Law Review* 611-688 at 646-48.

[34] Law Council of Australia, 'Human Rights and Technology' (Submission to the Australian Human Rights Commission in response to its 'Human Rights and Technology Issues Paper', 2018) at 14 https://lawcouncil.au/publicassets/85d805e0-14d8-e811-93fc-005056be13b5/3533%20-%20Human%20Rights%20and%20Technology.pdf.

[35] Carston Orwat, 'Algorithmic Discrimination From the Perspective of Human Dignity' (2024) 12 *Social Inclusion* https://doi.org/10.17645/si.7160; Felicity Bell, Lyria Bennett Moses, Michael Legg, Jacob Silove and Monika Zalnieriute, 'AI Decision-Making and the Courts: A Guide for Judges, Tribunal Members and Court Administrators' (Australiasian Institute on Judicial Administration, 2022) at 2.

[36] *Beian v. Romania* (no. 1) [2007] 30658/05 (6 December 2007) § 39; *Lupeni Greek Catholic Parish and Others v. Romania* [76943/11] (29 November 2016) § 116.

[37] *Donadze v. Georgia* [2006] 74644/01 (7 March 2006) § 35; *Kraska v. Switzerland* [1993] 13942/88 (19 April 1993) § 30; *Van de Hurk v. the Netherlands* [1994] 16034/90 (19 April 1994) § 59; *Perez v. France* [2004] 47287/99 (12 February 2004) § 80.

challenging questions. One concern here (also discussed earlier) is that data processing in AI systems can be impacted by bias in ways that perpetuate or create discriminatory outcomes.[38]

Another possible concern is the so-called 'relationality gap' between AI systems and humans. That is, AI knowledge is abstract and language-driven while human knowledge and interactions are rooted in a relational experience of the world.[39] This connects back to the classic syntax and semantic discussion as put forward by John Searle,[40] and since reconsidered in the context of symbol grounding problems and AI learning.[41] In the context of our discussion, these issues may provide justification for limitations on the use of generative AI in civil justice processes. For instance, in family law, the use of generative AI may require training data beyond the body of statute and case law to adjudicate cases that consider the sensitivity of matters where parties are particularly vulnerable.

A potential safeguard against concerns for discrimination or the relationality gap include human involvement and constant appraisal and refinement of programming in generative AI. There is additional potential benefit here of maintaining trust in the justice system and 'social legitimacy' of AI informed decisions by keeping humans in the loop.[42] However, it is also worth considering that even with human involvement, perceptions that AI generated content is objective and accurate may mean people might not appropriately scrutinise automated decisions.[43] Related to the idea of deference to generative AI on matters is a question about whether generative AI has legal authority comparable to that of a human judge, and therefore whether generative AI can act meaningfully within the obligations imposed by Article 6 as a judicial decision maker.[44]

2.5   Right to reasoned judgement

Article 6 of the ECHR imposes an obligation on courts and tribunals to provide sufficient reasons for their decisions, that is, 'adequately state the reasons on which they (judgments) are based'.[45] Such reasons demonstrate to the parties that their case has been actually 'heard'.[46] It also affords a party the opportunity to challenge and appeal a decision, with the possibility of having an appellate body review the decision.[47] The European Court of Human Rights has explained that: 'The right to be heard therefore includes not only the possibility to make submissions to the court, but also a corresponding duty of the court to show, in its reasoning, the reasons for which the relevant submissions were accepted or rejected'.[48] Nevertheless, this

---

[38] United Nations Educational, Scientific and Cultural Organisation, 'Recommendations on the Ethics of Artificial Intelligence' at 20; Council of Europe, 'European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment' at 8.

[39] Emily Bender, 'Resisting Dehumanisation in the Age of "AI"' (2024) 33(2) *Current Directions in Psychological Science* https://doi.org/10.1177/09637214231217286.

[40] John R Searle, 'Minds, brains, and programs' (1980) 3(3) *Behavioral and Brain Sciences* 417-424.

[41] See Steven Harnad, 'The Symbol Grounding Problem' (1990) 42 *Physica D: Nonlinear Phenomena* 335-346.

[42] Sofia Olhede and Patrick Wolfe, 'The growing ubiquity of algorithms in society: implications, impacts and innovations' (2018) 376(2128) *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 20170364 https://doi.org/10.1098/rsta.2017.0364.

[43] Elena Abrusci and Richard Mackenzie-Gray Scott, 'The questionable necessity of a new human right against being subject to automated decision-making' *International Journal of Law and Information Technology* (2023) 31(2) 114–143 https://doi.org/10.1093/ijlit/eaad013.

[44] Tania Sourdin, 'Judge v robot? Artificial intelligence and judicial decision-making' (2018) 41(4) *UNSW Law Journal* 1114.

[45] *Hirvisaari v Finland* (2001) ECHR 49684/99 (27 September 2001) § 30.

[46] *Fomin v Moldova* (2012) ECHR 36755/06 (11 January 2012) § 31.

[47] Ibid.

[48] Ibid.

does not require courts to give a detailed answer to every argument,[49] and the extent of the duty to give a reasoned decision vary according to the nature of the decision.[50] Overall, this guarantee to a reasoned decision ensures judicial transparency, protects individuals against arbitrariness, and for the administration of justice.

Beyond the potential for drafting judgments, generative AI applications may also be used by judges to support fact-finding. For instance, a Dutch judge has used ChatGPT to assist in estimating the lifespan of solar panels in a civil matter.[51] Unlike an expert witness or a specific information source, such a fact-finding exercise may be difficult to reconstruct for the purposes of transparency. In addition to accuracy concerns, it also poses challenges for the ability of litigants to respond to the 'facts' sourced from generative AI.

At the time of writing, the UK's guidelines on the judicial use of AI notes that generative AI does not necessarily provide convincing or accurate legal analysis or reasoning.[52] Similarly, the judicial guidelines in the state of Victoria, Australia make it clear that:

> Generative AI and Large Language Models create output that is not the product of reasoning. Nor are they a legal research tool. They use probability to predict a given sequence of words. Output is determined by the information provided to it and is not presumed to be correct.[53]

An area of scepticism relates to the ability of generative AI to participate in the kind of reasoning required of judicial decision making; that is, not merely deductive decisions based on facts and logic but involve discretionary decisions that include reflection on social values and the incorporation of 'soft decision-making factors' including intuition and inductive reasoning.[54] Due to the aforementioned relationality gap, the use of generative AI may well be less appropriate in decision making in some areas of civil disputes where more discretionary decisions are required, such as family law. By contrast, generative AI may be a tool for reasoning in commercial civil matters, where chronologies and decisions are often recorded in paper trails. An option here may be to 'apportion discrete types of judicial decision-making to human actors based on subject matter or legal issues which raise particular justice concerns'.[55]

A key concern is also a lack of transparency and explainability associated with the 'black box' characteristics of AI decision making, which can defy the idea of a 'reasoned decision'. This can invoke a transparency gap in which it may be impossible for humans to reconstruct and access the grounds on which AI came to a decision or output.[56] Explainability describes

---

[49] *Van de Hurk v. the Netherlands* [1994] ECHR 16034/90 (19 April 1994) § 61; *Case of Perez v. France* [2004] ECHR 47287/99 (12 February 2004) § 81.

[50] *Ruiz Torija v. Spain* [1994] ECHR 18390/91 (9 December 1994) § 29; *Hiro Balani v. Spain* [1994] ECHR 18064/91 (9 December 1994) § 27.

[51] See the civil law case of CLI:NL:RBGEL:2024:3636 (2024) § 5.7

[52] Courts and Tribunals Judiciary, 'Artificial Intelligence (AI) - Judicial Guidance' at 6.

[53] Principle 8 of the Supreme Court of Victoria, 'Guidelines for Litigants: Responsible Use of Artificial Intelligence in Litigation' (Supreme Court of Victoria, May 2014) www.supremecourt.vic.gov.au/forms-fees-and-services/forms-templates-and-guidelines/guideline-responsible-use-of-ai-in-litigation.

[54] Luís Greco, 'Judicial Power Without Judicial Responsibility: The Case Against Robot Judges', in Dário Moura Vicente, Rui Soares Pereira and Ana Alves Leal (eds), *Legal Aspects of Autonomous Systems. ICASL 2022. Data Science, Machine Intelligence, and Law* (vol 4., Springer, 2024) https://doi.org/10.1007/978-3-031-47946-5_12; Australian Law Reform Commission, 'Technology: what it means for Federal dispute resolution' (Issues Paper No 23., 1998) at 101.

[55] Tania Sourdin, 'Regulating Judge Artificial Intelligence (AI)', in Dário Moura Vicente, Rui Soares Pereira and Ana Alves Leal (eds), *Legal Aspects of Autonomous Systems. ICASL 2022. Data Science, Machine Intelligence, and Law* (vol 4., Springer, 2024) https://doi.org/10.1007/978-3-031-47946-5_12.

[56] Aziz Z. Huq, 'A Right to a Human Decision' at 636.

how much of AI reasoning can be accessible and is comprehensible if it is indeed accessible.[57] As captured by Greco's critique of 'robo-judges':

> For insofar as algorithms function as black boxes, they deny individuals any justification for the outcome adversely impacting them. Put plainly, individuals are denied the very thing that constitutes the lawfulness of the results affecting them, the thing that distinguishes the decision from a fiat pronouncement.[58]

To address the 'black box' problem, there have developments in explainable AI research which aims to create understandable AI models that can shed light on their decision-making processes.[59] Whether in fully automated systems or as support tools in judicial decision-making, the ability to explain the actions of generative AI tools is crucial for the right to a reasoned decision. There can be technical safeguards such as requiring the disclosure and independent verification of the underlying AI models, as well as procedural safeguards for litigants to be able to respond and challenge elements of the decision generated by AI.

Despite the safeguards and interpretations developed around Article 6, a crucial question remains unresolved in the face of advancing AI: does the right to a fair trial inherently encompass a right to a human judge? While Article 6 has been interpreted to guarantee various aspects of a fair trial, including access to a court, independence, impartiality, procedural fairness and the right to a reasoned decision, it has not explicitly addressed the nature of the decision-maker itself. As AI systems become increasingly sophisticated and their potential role in judicial processes expands, this gap in our interpretation of Article 6 becomes more pressing.

3   AI decision making in courts: the right to a human judge?

We have identified in Part 2 the safeguards in Article 6 which can be found explicitly listed in the article as well as extrapolated by the ECtHR. Unsurprisingly, the 'humanness' of the judge is not one of these safeguards. Over the past 75 years, it has not needed to be. Instead, Article 6 has been exercised in the face of threats which we already know about from centuries of human-driven justice, including inaccessible courts, corrupt judiciaries, opaque reasoning, and unfair procedures. We have shown that generative AI provides fresh challenges to these existing threats but may also provide some new opportunities to counter them.

Yet, the above analysis is incomplete, since generative AI, and AI more broadly, pose an additional threat to justice which requires considered analysis: the risk of dehumanisation in the context of AI judicial decision-making. We situate 'the right to a human judge' not as a novel idea to the corpus of human rights jurisprudence but advocate for a 'thick' reading of Article 6 which is underpinned by the principle of human dignity. In this Part, we go on to argue that this dignity-informed interpretation of Article 6 can be used to elucidate a right to a human judge implicit in Article 6. Dignity must also be a guiding principle in which all AI systems can be integrated into the justice system.

3.1   The threat of dehumanisation

First, we must establish what we mean by 'dehumanisation'. This concept is detected in various guises in the literature examining the 'datafication' of society, that is the abstraction of the

---

[57] Elena Abrusci and Richard Mackenzie-Gray Scott, 'The questionable necessity of a new human right against being subject to automated decision-making'.

[58] Luís Greco, 'Judicial Power Without Judicial Responsibility: The Case Against Robot Judges'.

[59] See chapter in this Handbook by Zijie Huang on 'Unleashing Creative Potential: Nurturing Trustworthy Generative AI'. See further Johannes Schneider, 'Explainable Generative AI (GenXAI): A Survey, Conceptualization, and Research Agenda' arXiv:2404.09554v1 [cs.AI] 15 Apr 2024 https://arxiv.org/html/2404.09554v1.

world and its phenomena—including humans and their characteristics, experiences, and behaviours—into quantifiable and analysable data.[60] Vociferous data acquisition by powerful actors has been described as 'echo[ing] imperialist modes of accumulation'.[61] The increasing access to and use of data by powerful actors, either state bodies or multinational corporations, has been criticised for its intrusive nature through surveillance and commodification of private human experience, which manifests as a form of dehumanisation.[62]

Dehumanisation also underlies discussions of the discriminatory impacts of data-driven algorithms, including the ways in which certain groups are treated, excluded, and marginalised. For example, in the case of predictive policing algorithms, studies have pointed out that biased data is created through patterns of concentrated police powers and surveillance on majority Black and racialised communities in the US. Such data leads to discriminatory outcomes that are then used to justify and perpetuate over-policing of those communities.[63]

As Deborah Raji points out, the dehumanisation occurs not only in the discrimination itself but also in the fiction of 'the data does not lie' and the purported accuracy or 'ground truth' in data sets that contain those biases.[64] Raji considers examples such as Google Image search results for 'healthy skin' that show only light-skinned women or ImageNet-trained models that label her (as a Black person) a 'bad person', a 'drug addict', or a 'failure'. She observes: 'data sets so specifically built in and for white spaces represent the constructed reality, not the natural one. To have accuracy calculated in the absence of my lived experience not only offends me, but also puts me in real danger'.[65]

The above examples point to another potential type of dehumanisation, connected to but distinct from discrimination: stripping the person of their individuality and autonomy. As Tennison has observed in the justice context, 'data by its nature flattens and categorises people, and isn't good at capturing complex contexts and nuance'.[66] If being treated humanely is to be treated as an individual, then people can be dehumanised in being not treated as individuals with autonomy and free will, but as categories. In other words, the use of generative AI in judicial decision making may abstract and objectify people into 'algorithmic categories'.[67]

The potential dehumanisation brought by the deployment of generative AI in the justice system is linked to the question of its 'understanding' of the outputs that it is producing or

---

[60] See the definition of Mayer-Schönberger and Cukier: 'to datafy a phenomenon is to put it in quantified form so that it can be tabulated and analyzed'. Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (London, UK: John Murray, 2013) p. 78.

[61] Jathan Sadowski, 'When data is capital: Datafication, accumulation, and extraction' (2019) 6(1) *Big Data & Society* https://doi.org/10.1177/2053951718820549, citing Jim Thatcher, David O'Sullivan, and Dillon Mahmoudi, 'Data colonialism through accumulation by dispossession: New metaphors for daily data' (2016) 34(6) *Environment and Planning D: Society and Space* 990–1006; Rosa Luxemburg, *The Accumulation of Capital (*New York, NY: Monthly Review Press, 1951).

[62] Jose van Dijck offers a critical interpretation of datafication as 'a means to *access [...]* and *monitor* people's behavior'. See Jose van Dijck, 'Datafication, Dataism and Dataveillance: Big Data Between Scientific Paradigm and Ideology' (2014) 12(2) *Surveillance & Society* 197–208 at 198. See also Shoshana Zuboff, *The Age of Surveillance Capitalism* (London, UK: Profile Books, 2019).

[63] Rashida Richardson, Jason Schultz and Kate Crawford, 'Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice' (2019) 94 *New York University Law Review* 192-233.

[64] Deborah Raji, 'How our data encodes systematic racism', MIT Technology Review, 10 December 2020 www.technologyreview.com/2020/12/10/1013617/racism-data-science-artificial-intelligence-ai-opinion.

[65] Ibid.

[66] Jenni Tennison, 'Future of Courts' (Panellist remarks given at an event sponsored by The Nuffield Foundation and Legal Education Foundation in collaboration with UCL Faculty of Laws, Connected by Data, 14 May 2024) https://connectedbydata.org/events/2024-05-14-future-of-courts.

[67] Yuval Shany, 'The Case for a New Right to a Human Decision Under International Human Rights Law' (October 4 2023) http://dx.doi.org/10.2139/ssrn.4592244 at 38; Elena Abrusci and Richard Mackenzie-Gray Scott, 'The questionable necessity of a new human right against being subject to automated decision-making'.

approaching something like 'reason'.[68] It has been stressed that LLMs do *not* exhibit natural language understanding.[69] However, coherence is in the eye of the beholder. While LLMs may not 'understand' what the sequences of words mean, nor have any intent to communicate, that does not stop humans' predisposition to interpret meaning and intent *into* what they listen to or read.[70] Anthropomorphising the output risks lulling the user into forgetting they are the only one in the conversation who is imbuing meaning. A potential aspect of dehumanisation could arise here, from the inherent deception of the interaction, should the human interpret coherence into the machine which has none.

In the context of judicial decision making, social-legal and psychological research on procedural justice offers valuable insights into how individuals experience and perceive judicial processes. An individual's 'satisfaction' with their experience of court is not merely determined by the verdict or outcome of their case. Instead, it is more directly linked to their perceptions of how fair the *process* itself was.[71] This insight is consistent across age, gender, education level, income, ideological beliefs, or ethnicity. In contrast, ideas of what constitutes a fair *outcome* in any given case can vary widely between different demographic groups.[72] Furthermore, court users' perceptions of fairness are shown to impact the legitimacy attributed to the institutions involved, and the likelihood of people to accept outcomes that are unfavourable to them.[73] In this context, Tyler identifies four factors that are important to individuals when they are subject to decision making by an authority, and what it means to be treated fairly and 'humanely':

(i) the opportunity for individuals to express their views during the decision-making process (participation or voice),
(ii) unbiased decision-making (neutrality),
(iii) being treated with respect (respect), and
(iv) are willing to help and support them (trustworthiness).[74]

Considering these factors, we identify below some potential risks of dehumanisation should generative AI increasingly replace the functions of a human judicial decision maker.

3.1.1 Voice

First, the concept of participation, or 'voice', is crucial to one's perception of fairness in judicial proceedings. It allows individuals to feel heard and to contribute their perspective to the decision-making process. However, the use of generative AI in this context presents significant

---

[68] Jacob Browning and Yann LeCun, 'AI and the limits of language', Noema Magazine, 23 August 2022 www.noemamag.com/ai-and-the-limits-of-language//.

[69] Emily Bender and Alexander Koller, 'Climbing towards NLU: On meaning, form, and understanding in the age of data' (Proceedings of the 58th annual meeting of the association for computational linguistics, Association for Computational Linguistics Anthology, July 2020) https://10.18653/v1/2020.acl-main.463.

[70] Ibid. See further Bender, Gebru, McMillan-Major, & Shmitchell who coined the term 'stochastic parrots' to describe the way in which LLMs 'haphazardly [stitch] together sequences of linguistic forms… according to probabilistic information about how they combine, but without any reference to meaning'. Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?' (Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 1 March 2021) https://doi.org/10.1145/3442188.3445922. More generally, see philosophical discussion on this in John R Seale, 'Minds, Machines and Programs'.

[71] Tom Tyler, 'Introduction' in *Procedural justice: Volume I. Procedural justice and the dynamics of authority* (ed. Tom Tyler), Vol. 1, pp. xi-xxiii (Burlington, V.T: Ashgate, 2005), pp. xii-xiii.

[72] Ibid, xv.

[73] Ibid, xii; John Thibaut and Laurens Walker, *Procedural Justice: A Psychological Analysis Hillsdale* (N.J. Lawrence Erlbaum Associates, 1975).

[74] Steven Blader and Tom Tyler, 'A Four-component Model of Procedural Justice: Defining the Meaning of a "Fair" Process', (2003) 29 *Personality and Social Psychology Bulletin* 747–758, 748–749; Tom Tyler, 'Procedural Justice and the Courts', (2007) 44 *Court Review* 26–31.

challenges. While generative AI can process and analyse input, generating human-like responses, it cannot genuinely 'understand' in the way a human judge can. The output, while seemingly natural, is generated based on patterns in the training data rather than genuine understanding. The richness of tone, emotion, context, and cultural nuances that humans naturally perceive and interpret are processed by generative AI without true comprehension. This fundamental limitation of generative AI systems undermines the essence of participation.

Moreover, generative AI may give an illusion of a seemingly meaningful dialogue, asking follow-up questions and exploring various aspects of a case. However, this interaction is fundamentally different from a conversation with a human judge. The AI's responses, while potentially relevant and articulate, are generated based on statistical patterns from processing large amounts of data and executing complex algorithms. Without true understanding or empathy, the application of generative AI in judicial processes can lead to a false sense of the individual being heard. This can potentially be more dehumanising than no interaction at all.

An additional reflection emerges from potential differences in cultural communication and expression styles across Europe. There is an emerging body of research that identifies the problem of bias in relation to how generative AI outputs around non-Western cultural contexts entail 'superimposed Western interpretations of those contexts'.[75] Such models may struggle to accurately respond to diverse cultural contexts, potentially leading to misunderstandings or misrepresentations in communications with individuals from non-Western communities.

3.1.2 Neutrality

Large language models are trained on vast datasets of human-generated content. These datasets inevitably contain biases present in society, which the AI may amplify in its outputs. Unlike human judges who can be trained to recognise and mitigate their biases, generative AI systems may perpetuate and even exacerbate biases without the ability to self-reflect or adjust based on changing and nuanced ethical considerations.

The concept of neutrality in judicial decision making goes beyond mere absence of bias. It requires a nuanced understanding of context and an ability to recognise and account for one's own limitations. There is a common misconception that empathy is at odds with impartiality. However, empathy is better understood as 'our capacity to better comprehend—through both knowledge and feeling—another's perspective by trying to view the world from that person's position, rather than simply observing another's position from where we stand'.[76] Sourdin highlights the underlying importance of the emotional authority of judges in moral decision making which allows judges to respond to parties in ways that uphold compassion and underscore fundamental rights of dignity.[77] While generative AI can produce empathetic-sounding language, it is not capable of understanding the lived experiences that shape an individual's actions and circumstances.

Building on this, we argue that neutrality in judicial decisions require a deep understanding of societal norms, cultural contexts and evolving social values. Generative AI is unable to achieve this level of understanding of changing social changes and emerging ethical

---

[75] Dan M Kotliar, 'Data orientalism: on the algorithmic construction of the non-Western other' (2020) 49(5) *Theory and Society* 919. See further Sourojit Ghosh, Pranav Narayanan Venkit, Sanjana Gautam, Shomir Wilson and Aylin Caliskan, 'Do Generative AI Models Output Harm while Representing Non-Western Cultures: Evidence from A Community-Centered Approach' (paper accepted at 7th AAAI/ACM Conference on AI, Ethics, and Society, 21-23 October 2024) https://arxiv.org/pdf/2407.14779.

[76] Rebecca K Lee, 'Judging Judges: Empathy as the Litmus Test for Impartiality' (2014) 82 *University of Cincinnati Law Review* 145-206. On the importance of humility to the role of the judge, see also Sir Mark Hedley, *The Modern Judge: Power, Responsibility and Society's Expectations* (Jordan Publishing, 2016).

[77] Tania Sourdin, 'Regulating Judge Artificial Intelligence (AI)', in Dário Moura Vicente, Rui Soares Pereira and Ana Alves Leal (eds), *Legal Aspects of Autonomous Systems. ICASL 2022. Data Science, Machine Intelligence, and Law* (vol 4., Springer, 2024) https://doi.org/10.1007/978-3-031-47946-5_12.

considerations. This can result in decisions that fail to account for the current social and human context.

3.1.3 Respect

The element of treating an individual with respect in judicial proceedings encompasses more than just polite treatment. It involves recognising the inherent dignity of every person and acknowledging their rights and humanity throughout the process. The content (output) from generative AI tools may be capable of conveying some acknowledgement of an individual's circumstances and experiences and their rights, but it cannot go further. This lack of genuine recognition can be particularly damaging in cases involving sensitive or traumatic experiences, where human understanding and compassion are crucial.

Human judges can adapt their communication style based on the needs and capabilities of the individuals appearing before them. Judicial styles can vary, e.g., some judges may be more 'responsive' than others and/or 'show more emotion and compassion or be oriented towards therapeutic justice'.[78] Nevertheless, the absence of personal interactions with a human judge presents a risk of 'dehumanising' communication. Generative AI, despite the ability to produce varied language, lacks the intuitive ability to truly adapt its communication style to the emotional and cognitive needs of diverse individuals. The technology is, after all, selecting the statistically more likely word among different options. For example, it may end up using overly technical language when describing personal and emotional matters, which can contribute to a sense of being treated as a 'case' rather than a person.

3.1.4 Trustworthiness

Trustworthiness involves not just the reliability of decisions, but also the perception that the judiciary is able to support those who are subject to its authority. The introduction of generative AI into judicial decision making poses several challenges to building and maintaining this trust. While generative AI may provide 'explanations' for its decisions, these explanations are generated post-hoc and may not reflect the actual pathways by which the decision was made, which can potentially eroding users' trust in the process. The complex neural networks behind advanced AI models are often opaque, even to their creators. It is unsurprising that explainable AI has become a significant area of research among legal AI researchers.[79] Their 'black box' characteristics make it difficult to truly understand how and why certain decisions were made. AI researchers have further pointed out the significant challenges of testing and evaluating LLMs due to the 'sheer scale of data and training, plus the proprietary nature of LLMs'.[80]

Where there are inaccuracies produced by generative AI, such errors in a judicial context could have severe consequences and significantly undermine trust in the system. The potential for these errors, even if they are rare, can make it difficult for individuals to fully trust decisions flowing from generative AI-driven judicial processes.

Moreover, AI systems cannot be held accountable for their decisions in the way human judges can. Unlike human judges who can be questioned, challenged, and held personally accountable for their decisions and reasoning, it is doubtful that AI systems possess moral

---

[78] Tania Sourdin, 'Judge v robot? Artificial intelligence and judicial decision-making'.

[79] See, for example, Joe Collenette, Katie Atkinson and Trevor Bench-Capon, 'Explainable AI tools for legal reasoning about cases: A study on the European Court of Human Rights' (2023) 317 *Artificial Intelligence* 103861; Ashley Deeks, 'The Judicial Demand for Explainable Artificial Intelligence' (2019) 119 *Columbia Law Review* 1829.

[80] Ananya, 'AI's Understanding and reasoning skills can't be assessed by current tests', Science News, 10 July 2024 www.sciencenews.org/article/ai-understanding-reasoning-skill-assess.

culpability.[81] This diffusion of responsibility can further erode users' trust in the judicial system. The use of generative AI here requires serious rethinking of judicial accountability and responsibility. When problems arise in the process, who (or what) do we attribute responsibility?[82] As Greco highlights, this issue is not one to be taken lightly, especially considering the potential use of generative AI in decision making may represent a 'historically unprecedented detachment of the exercise of power from individual responsibility'.[83]

Overall, the potential dehumanising effects of the use of AI in judicial decision making are multifaceted and can significantly undermine the right to a fair trial. Considering this threat, we argue that the principle of dignity should underpin a 'thick' interpretation of the protections offered by Article 6.

3.2   Protection from dehumanisation: human dignity and Art 6 of the ECHR

The concept of 'human dignity' became prevalent in international human rights legal instruments after the atrocities of the second world war.[84] The Preamble of the UN Charter in 1945 cites the 'untold sorrow to mankind' of the World Wars alongside its statement of purpose: to 'reaffirm faith in fundamental human rights, in the dignity and worth of the human person'.[85] Similarly, the Universal Declaration of Human Rights (UNHR) in 1948 recognises 'barbarous acts which have outraged the conscience of mankind'[86] in its Preamble before asserting the 'inherent dignity and the equal and unalienable rights of all members of the human family'.[87] It also goes further than the UN Charter and dedicates its first Article to dignity: 'All human beings are born free and equal in dignity and rights. They are endowed with reason'.

Human dignity maintains its prevalence in human rights instruments throughout the second half of the twentieth century. The 1976 International Covenant on Civil and Political Rights states that rights 'derive from the inherent dignity of the human person', and that the recognition of 'inherent dignity' and of 'the equal and inalienable rights of all members of the human family' is the 'foundation of freedom, justice and peace in the world'.[88] Further, in more specific international human rights legal instruments, for example those aimed at children or people with disabilities,[89] human dignity remains a ubiquitous concept.

---

[81] It is outside the scope of this chapter to discuss the philosophical debates on whether autonomous AI has moral responsibility. A good discussion of relevant issues can be found in Mihaela Constantinescu, Constantin Vică, Radu Uszkai and Cristina Voinea, 'Blame It on the AI? On the Moral Responsibility of Artificial Moral Advisors' (2022) 35 *Philosophy and Technology* 1-26 https://doi.org/10.1007/s13347-022-00529-z.

[82] For more discussion, see Luís Greco, 'Judicial Power Without Judicial Responsibility: The Case Against Robot Judges'.

[83] Ibid.

[84] Christopher McCrudden, 'Human Dignity and Judicial Interpretation of Human Rights' (2008) 19(4) *European Journal of International Law* 655–724 at 667.

[85] Paragraph 2 in the Preamble of the Charter of the United Nations, 1 UNTS XVI, 24 October 1945.

[86] Paragraph 2 of the Universal Declaration of Human Rights, 217 A (III), 10 December 1948.

[87] Ibid Paragraph 3.

[88] Paragraphs 2 and 3 of the Preamble of the International Covenant on Civil and Political Rights, United Nations, Treaty Series, vol. 999, p. 171, 16 December 1966.

[89] Sebastian Heselhaus and Ralph Hemsley, 'Human Dignity and the European Convention on Human Rights', in Paolo Becchi and Klaus Mathis (eds), *Handbook of Human Dignity in Europe* (Springer, 2019) pp. 1-24.

While it was unusual for the drafters of the ECHR not to cite human dignity explicitly in 1950,[90] a careful reading of the ECHR reveals its implicit presence.[91] The Preamble directly invokes the UDHR. The UDHR was only a few months old when the ECHR was being drafted. Indeed, many of the ECHR drafters had also drafted the UDHR, and the ECHR was in many ways a *continuation* of their work, not a separate project. The Preamble of the ECHR goes on to state its purpose not to rewrite or replace the UNDR, but rather to 'take the first steps for the collective enforcement of certain of the rights stated in the Universal Declaration'. The UNDR is therefore not superseded but rather incorporated as the foundation of the ECHR, the latter seeking make the human dignity and rights in the former practically enforceable. The ECtHR has repeatedly agreed with this view and found that 'human dignity and human freedom' are the 'very essence' of the ECHR.[92]

While human dignity is relevant to the ECHR, how to apply the principle can sometimes be unclear. McCrudden has argued that dignity as a standalone right is too vague.[93] The fact that the drafters of the ECHR did not explicitly include dignity as a standalone right may reinforce this view. Instead, it is more useful to consider it as a fundamental principle underlying the interpretation of specific rights within the ECHR. In this sense, the principle of dignity can support the interpretation of other rights, including their scope and application, having what has been called an 'indirect legal effect'.[94]

We consider this function of human dignity to be particularly apt to the ECHR when understanding it as a living instrument. As Sir Humphrey Waldock (former President of the Court) once put it: 'The meaning and content of the provisions of the ECHR will be understood as intended to evolve in response to changes in legal or social concepts.'[95] This approach is particularly valuable when faced with novel contexts, such as the integration of AI into judicial processes. Human dignity can serve as a guiding principle, ensuring that the evolution of rights interpretation remains consistent with the ECHR's core values.

In delving deeper into the interpretative value of the principle, McCrudden has identified a minimum core of three elements making up a substantive idea of dignity. The first, the 'ontological claim', is that every human being possesses an intrinsic worth, merely by being human. The second is that this intrinsic worth should be recognised and respected by others, and some forms of treatment by others are inconsistent with, or required by, respect for this intrinsic worth. This is the 'relational claim'. The third is the recognition that 'the state should

---

[90] In fact, the word 'dignity' was included for the first time in the Convention in 2002 upon the introduction of Protocol 13 on the abolition of the death penalty. The Preamble states at Paragraph 1: 'everyone's right to life is a basic value in a democratic society and that the abolition of the death penalty is essential for the protection of this right and for the full recognition of the inherent dignity of all human beings'.

[91] Jean-Paul Costa has also argued that the absence of human dignity is due to technical and practical alignments of the ECHR rather than its irrelevance to the ideas that underpinned it. See Jean-Paul Costa, 'Human Dignity in the Jurisprudence of the European Court of Human Rights', in Christopher McCrudden (ed.), *Understanding Human Dignity - Proceedings of the British Academy* (vol. 192, Oxford University Press, 2013) pp. 393-402, p. 394.

[92] *Christine Goodwin v UK* [2002] 28967/95 (11 July 2002) § 90.

[93] Christopher McCrudden, 'Human Dignity and Judicial Interpretation of Human Rights' (2008) 19(4) *European Journal of International Law* 655–724.

[94] Sebastian Heselhaus and Ralph Hemsley, 'Human Dignity and the European Convention on Human Rights', in Paolo Becchi and Klaus Mathis (eds), *Handbook of Human Dignity in Europe* (Springer, 2019) pp. 1-24, quoting Lennart von Schwichow, 'Die Menschenwürde in der EMRK', in W Marauhn (ed) Jus Internationale et Europaeum (vol 123, 2016).

[95] Sir Humphrey Waldock, 'The Evolution of Human Rights Concepts and the Application of the European Convention on Human Rights', in André Gros and Paul Reuter, *Mélanges offerts à Paul Reuter: le droit international: unité et diversité* (Paris: Pedone, 1981) p. 547.

be seen to exist for the sake of the individual human being, and not vice versa (the limited-state claim)'.[96]

While McCrudden himself still finds dignity to have an 'impossible vagueness' in the diversity of contextual meanings that it can take on, he agrees with Clapham that there are four values with which dignity seems particularly closely and consistently related to: (1) the prohibition of all types of inhuman treatment, humiliation, or degradation by one person over another; (2) the assurance of the possibility for individual choice and the conditions for each individual's self-fulfilment autonomy, or self-realisation; (3) the recognition that the protection of group identity and culture may be essential for the protection of personal dignity; (4) the creation of the necessary conditions for each individual to have their essential needs satisfied.[97]

Given these potential threats to human dignity, we propose a 'thick' interpretation of Article 6 of the ECHR, guided by the principle of human dignity. This interpretation would: (1) recognise the intrinsic worth of everyone appearing before a court, ensuring that they are treated as unique human beings rather than mere data points; (2) guarantee meaningful participation in the judicial process, including the ability to be truly heard and understood; (3) ensure that decisions are made with genuine understanding of the human context, including cultural and personal factors; (4) maintain transparency and explainability in the decision-making process, allowing individuals to understand and potentially challenge the reasoning behind judgments; and, (5) preserve moral accountability in the judicial system, recognising that justice is fundamentally a human concept that requires human understanding and responsibility.

Based on these considerations, we argue that a dignity-informed reading of Article 6 implicitly secures the right to a human judge. This interpretation recognises that the essence of a fair trial, as envisioned in the ECHR, relies on human judgment, empathy, and understanding that cannot be fully replicated by AI systems.

3.3   Potential counterarguments

While the case for a human judge based on dignity is strong, several counterarguments merit consideration. First, it may be argued that respecting human dignity includes respecting an individual's choice to have their case decided by an AI if they prefer. Second, as discussed earlier in Part 2, the current human-based judicial system often results in significant delays and unequal access to courts. An AI-based system, especially using generative AI, could potentially offer quicker, more affordable and consistent access to justice. Finally, some would argue that human judges are susceptible to various biases, which can compromise the fairness of trials. AI systems, if properly designed, could potentially reduce certain types of bias, thereby enhancing the dignity of all parties by ensuring more equitable treatment.

While these counterarguments raise valid points, they do not necessarily negate the need for human judges. Instead, they highlight the need for a nuanced approach that balances various aspects of human dignity. We propose the following considerations. First, rather than replacing human judges with AI, efforts should focus on addressing and mitigating human biases through training, diverse judicial appointments, and accountability measures. Second, AI can be valuable in automating non-judicial tasks, providing research assistance, and identifying potential biases or inconsistencies for human review. This approach preserves human judgment while leveraging AI's benefits. Third, for cases where parties might prefer a more automated or rapid process, improving access to ADR mechanisms (where informed consent can be obtained) could provide options without compromising the core judicial system.

---

[96] Christopher McCrudden, 'Human Dignity and Judicial Interpretation of Human Rights'.
[97] Ibid at 686, citing Andrew Clapham, *Human Rights Obligations of Non-State Actors* (Oxford University Press, 2006) at 545–546.

We note that the AI Act refers to the use of AI in ADR, with Recital 61 providing that 'AI systems intended to be used by alternative dispute resolution bodies for [the purposes of the administration of justice and democratic processes] should also be considered to be high-risk when the outcomes of the alternative dispute resolution proceedings produce legal effects for the parties'. Fourth, whether AI is used in supporting roles or in ADR, emphasis should be placed on transparency and explainability of AI-assisted processes, ensuring that human dignity is respected through the informed participation of all stakeholders. Finally, the integration of AI into any part of the judicial system should be subject to ongoing ethical review, with a focus on preserving human dignity in upholding the right to a fair trial.

Dignity is particularly important in the context of interactions between the civil justice system and marginalised and vulnerable populations. For instance, the European Commission Strategy for the Rights of Persons with Disabilities 2021-2030 and the Directive (EU) 2019/882 on the accessibility requirements for products and services specifically emphasise the importance of accessibility of use of technology for people living with disabilities or functional limitations.[98] In this context, the use of generative AI in courts or in parties' preparation for judicial proceedings must consider the accommodation and support of needs associated with the disabilities that people may live with. Here assistive AI technology, designed to support people with specific disabilities to promote dignity, independence, and facilitate participation, may be used to support their access to courts.[99] The deployment of generative AI in this context would recognise the importance of human interaction (including interaction with a human judge) in facilitating access and meeting support access to justice needs of marginalised groups in an inclusive way, which is underpinned by respect for dignity of all.

4   Conclusion

The United Nations Educational, Scientific and Cultural Organization's Recommendations on the Ethics of Artificial Intelligence reiterates that in interactions with AI, humans 'should never be objectified nor should their dignity be otherwise undermined, or human rights and fundamental freedoms violated'. [100] The principle of human dignity, deeply rooted in international human rights law and implicitly present in the ECHR, provides a crucial guiding framework for interpreting the right to a fair trial under Article 6 of the ECHR in the face of technological advancements. This necessarily implies a right to have one's case heard and decided by a human judge. While generative AI presents significant opportunities to enhance certain aspects of the right to a fair trial, its implementation must be carefully managed to avoid the risks of dehumanisation. A dignity-centred interpretation of Article 6 strongly suggests the need to preserve human judges at the core of the judicial process in the age of AI.

---

[98] See Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services, OJ L 151.

[99] Mihalis Kritikos, 'Assistive technologies for people with disabilities: Part IV: Legal and socio-ethical perspectives' (Scientific Foresight Unit (Science and Technology Options Assessment) of the European Parliamentary Research Service, 2018) at 4 www.europarl.europa.eu/RegData/etudes/IDAN/2018/603218/EPRS_IDA(2018)603218(ANN4)_EN.pdf.

[100] United Nations Educational, Scientific and Cultural Organisation, 'Recommendations on the Ethics of Artificial Intelligence' at 18.