

# AI in our Justice System

A JUSTICE report led by Sophia Adams Bhatti January 2025





# Contents

---

<b>Contents</b> .....	<b>3</b>	GOAL TWO: Fair and lawful decision-making .....	<b>27</b>
<b>Foreword</b> .....	<b>4</b>	i. How this goal upholds human rights and the rule of law .....	<b>27</b>
<b>Executive Summary</b> .....	<b>5</b>	GOAL THREE: Scrutiny and public engagement .....	<b>29</b>
<b>Our response</b> .....	<b>5</b>	i. How this goal upholds human rights and the rule of law .....	<b>29</b>
<b>1. BACKGROUND: What are we talking about?</b> .....	<b>7</b>	ii. Opportunities .....	<b>30</b>
a. What is “AI”? .....	<b>8</b>	<b>4. MANAGING RISKS: Innovating with AI to support the justice system....</b>	<b>31</b>
b. Who and what is our “justice system”? .....	<b>10</b>	a. Goal-led .....	<b>32</b>
c. What is “Justice AI”? .....	<b>11</b>	b. Duty to act responsibly .....	<b>32</b>
d. Who is responsible? .....	<b>13</b>	c. Understanding the Risks .....	<b>33</b>
<b>2. CONTEXT: Why are we talking about it?</b> .....	<b>15</b>	i. <i>Technical risk factors</i> .....	<b>33</b>
a. The stakes: our justice system and the rule of law .....	<b>16</b>	ii. <i>Individual harm risk factors</i> .....	<b>35</b>
b. The innovation context .....	<b>18</b>	iii. Systematic risk factors .....	<b>38</b>
c. Our approach .....	<b>20</b>	d. Risk mitigation .....	<b>45</b>
<b>3. ANALYSIS: Goals and Opportunities for the use of AI in our justice system</b> .....	<b>24</b>	<b>5. CONCLUSIONS</b> .....	<b>47</b>
GOAL ONE: Equal and effective access to justice .....	<b>24</b>	<b>6. ANNEX 1: Illustrative Framework Assessment Tool</b> .....	<b>46</b>
i. How this goal upholds human rights and the rule of law .....	<b>25</b>	<b>Acknowledgements</b> .....	<b>50</b>
ii. Opportunities .....	<b>25</b>		

# Foreword

---



Our justice system is both simultaneously a keystone in our democracy and in peril. A system which delivers justice with equality and fairness, and effectively, is crucial for society and the economy alike. The inability to rely on the justice system, or the weakening of its core tenants, undermines not just the justice system, but the very fabric of our society.

AI, or indeed the promise of AI yet to come, holds the potential to solve many genuine problems of people and planet – including some of the issues we are grappling with in the justice sector. As the Prime Minister announced recently AI is “a force for change that will transform the lives of working people for the better”.

It equally has the potential, as we have already seen, to cause significant harms. This does not mean we close the door to the many opportunities it offers, but rather open it with clear expectations of what good looks like, what outcomes we are seeking, the risks we are willing to take as society, and the red lines we want to put in place.

In the Prime Minister’s own words, “Government has a responsibility to make it work for working people”, but I would go further, and believe

that all those involved in the development of AI have a responsibility to safeguard fundamental human rights.

Questions of risks, red lines, outcomes sought - these are not easy questions to answer, but that is not an excuse to simply hope and pray that it all works out ok in the end. Applying our minds and taking deliberate steps to shape the future we wish for is a necessity if as a society we are to reap the benefits of AI, enable our public services to flourish and avoid the pitfalls – obvious and less so.

Some of the action needed will require boldness of vision, and an eye on the longer term. Given the state of current need, political pressures and the near-term problems it is easy to see how decision makers, both in the public and private sectors, might choose to put off until tomorrow some of these concerns. But to do so would be a mistake.

The opportunity is now – this report is not an anti-innovation agenda, far from it, rather it is a call to action to make a commitment to use AI to support the strengthening of our justice system and do so by reaffirming our commitments to the rule of law and human rights. Responsibility to invest in the future we will all eventually bequeath to generations to come rests not only with government ministers and policy makers, but with all those involved in the supply chain of AI in the justice system – from developers, through to purchasers and users whether they be in public bodies or private companies. The stakes are high and to get it right requires all our collective contributions and effort.

I hope that this report is a helpful tool to help pave the way for meaningful and deliberate use of AI which strengthens the rule of law and in turn our democracy.

# Executive Summary

## How can we use AI to improve people's lives?

This is a question for all of society. In answering it, we cannot ignore the problems which currently beset our justice system. From barriers to understanding the law and getting trustworthy legal advice, to court backlogs and overcrowded prisons, there are substantial problems across our justice system. These problems mar the lives of those affected and undermine the trust and confidence of wider society in the rule of law.

AI is not the single solution to the problems faced by the system, but it is clear that it is now part of the toolkit. Any new addition to that toolkit will inevitably be considered for its innovative potential to tackle old and new challenges. Although the promise of efficiency and effectiveness is tempting, and ought not be ignored, a critical eye is crucial.

The stakes are high; the justice system holds people's lives in its hands and plays a vital role in democratic societies. Delivering justice well is complex; it needs investment, and the capability to keep pace with changing societal needs. In this context, AI has potential, if deployed well, to be of great service to the strengthening of our justice system. But AI is far from a silver bullet: it can lack transparency, embed or exacerbate societal biases, and can produce inaccurate outputs which are nevertheless convincing to the people around it.

### *a. Our response*

**People need a justice system which they can trust to uphold the rule of law and protect human rights of everyone.** This means that attempts to improve the system through reforms and innovations, should have the core tenants of the rule of law and human rights embedded in their strategy, policy, design and development.

This paper proposes a framework to achieve this in the context of innovating with AI, consisting of two requirements:

#### **1. Goal-led: Have a clear objective of improving one or more of the core fundamental goals of a well-functioning justice system, which we synthesise as:**

- Equal and effective access to justice
- Fair and lawful decision-making
- Openness to scrutiny

Being goal-led guards against opportunism and ill designed changes which, although aimed at efficiency, drive risks which weaken the system. It ensures that innovations are targeted at genuine use cases which can help deliver better outcomes.

**2. Duty to act responsibly: All those involved in the deployment of AI within the justice system have a responsibility during the design, development and deployment of AI to ensure that the core features of the rule of law and human rights are embedded at each stage.**

This should include identifying risks and interrogating their impact, to prevent future harms. Furthermore, there should be an obligation to pause, rethink, redesign or even stop development or deployment if significant risks to the rule of law or human rights are identified. The degree of expertise and understanding of human rights and the rule of law will naturally differ across the ‘supply chain’ of AI. Some - such as the Ministry of Justice - will have in-depth experience and knowledge, whereas others, such as those in the tech field may have limited experience. It is the responsibility of those actors with greater knowledge and understanding of human rights and the rule of law to ensure that they clearly set expectations and boundaries for the other less experienced actors. However, this is not to let those without a legal background (for example in the tech industry) off the hook – each actor in the ‘supply chain’ must act responsibly, and in doing so this results in a stronger overall outcome.

This framework is deliberately simple. Clarity of focus on the purpose of the justice system allows for a clear line of sight between the many potential uses of AI and those which are genuinely in service of a justice system that upholds the rule of law and protects human rights.

This report is for all those who have the power to shape our justice system: legal practitioners, policymakers and developers in different contexts across the system, who may be considering innovation with AI. It also serves a second purpose for wider society, who may use it to scrutinise how others are innovating with AI in the justice system.

As a macro framework it applies equally to all areas of the justice systems - from criminal, to civil, to corporate litigation through to family. However, we recognise that in each of these areas the detailed considerations will be different, and the methods by which risks are managed will need specific approaches. As such we intend to further develop the practical application of this framework for each specific area and welcome contributions to inform the next stage of our work.





# 1. Background: What are we talking about?



## a. What is “AI”?

There is no definitive definition of artificial intelligence, or “AI”, although it is useful to set out some prominent explanations and some of the key terminology. As we explain later in the report, knowledge and understanding of AI plays a crucial role in being able to ensure that its use benefits the justice system, rather than harming it.

The OECD’s definition, updated May 2024, reads:

“An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.”<sup>1</sup>

The definition used in the UK Government Data Ethics Framework is as follows:

“AI can be defined as the use of digital technology to create systems capable of performing tasks commonly thought to require intelligence. AI is constantly evolving, but generally it:

- involves machines using statistics to find patterns in large amounts of data
- is the ability to perform repetitive tasks with data without the need for constant human guidance.”<sup>2</sup>

However, the UK Government added “autonomy” and “adaptiveness” to the above in 2023 (in common with the updated OECD definition). These two factors are important in understanding the functional capabilities of AI:

***“The ‘adaptivity’ of AI can make it difficult to explain the intent or logic of the system’s outcomes: AI systems are ‘trained’ – once or continually – and operate by inferring patterns and connections in data which are often not easily discernible to humans. Through such training, AI systems often develop the ability to perform new forms of inference not directly envisioned by their human programmers. The ‘autonomy’ of AI can make it difficult to assign responsibility for outcomes: Some AI systems can make decisions without the express intent or ongoing control of a human.”<sup>3</sup>***

This report takes a broad understanding of AI, to include various levels of autonomy and adaptiveness, according to the OECD definition. This ranges from the less autonomous and adaptive AI – such as automated processes which use rules-based algorithms – to the more autonomous and adaptive AI – inclusive of generative AI and deep learning technology.



<sup>1</sup> OECD, [Recommendation of the Council on Artificial Intelligence](#) (2019); OECD, [Explanatory memorandum on the updated OECD definition of an AI system](#) (2023).

<sup>2</sup> Central Digital & Data Office, [Data Ethics Framework: glossary and methodology](#) (2020); Department for Science, Innovation and Technology, Office for Artificial Intelligence and Centre for Data Ethics and Innovation, [A guide to using artificial intelligence in the public sector](#) (2019).

<sup>3</sup> Department for Science, Innovation and Technology and Office for Artificial Intelligence, [A pro-innovation approach to AI regulation](#) (2023).



To aid users of this framework, we have set out some of the key terminology.

<b>Algorithm</b>	“[A] set of mathematical instructions or rules ... given to a computer ... to calculate an answer to a problem” <sup>4</sup>
<b>Artificial general intelligence (AGI)</b>	A system that can understand, learn, and apply intelligence equivalent to or above human intelligence across a broad range of fields. Some, such as OpenAI and IBM, use the term <b>AGI</b> to explicitly describe theoretical systems that do not exist yet. Others, such as Google and the ISO, do not explicitly caveat that currently an <b>AGI</b> system is theoretical. <sup>5</sup>
<b>Automated decision-making</b>	The UK GDPR defines <b>automated decision-making</b> as “a decision based solely on automated processing, including profiling.” These decisions produce legal or similarly significant effects concerning a person or a decision of similar effect, such as a decision to grant a loan. <sup>6</sup>
<b>Automation bias</b>	A tendency by humans to overly rely on automated processes and decision-making over human judgment. <sup>7</sup>
<b>Data</b>	The “information in an electronic form that can be stored and processed by a computer” <sup>8</sup>
<b>Deep learning</b>	A type of <b>machine learning</b> , which can make inferences and connections from data while unsupervised due to having many more layers of neural networks than traditional <b>machine learning models</b> (between 3 and thousands of layers). <sup>9</sup>
<b>Deterministic model</b>	A <b>model</b> that always produces the same output if given the same input. <sup>10</sup>
<b>Foundation model</b>	A <b>model</b> “capable of range of general tasks (such as text synthesis, imagine manipulation and audio generation),” e.g., ChatGPT. <sup>11</sup>
<b>General-purpose AI (GPAI)</b>	See <b>foundation model</b> . Not to be confused with <b>artificial general intelligence (AGI)</b> .
<b>Generative AI (GenAI)</b>	A <b>model</b> that “can create (‘generate’) content that is ... complex, coherent, [and] original. For example, a generative AI model can create sophisticated essays or images.” <sup>12</sup>
<b>Hallucination</b>	The production of false or inaccurate information by <b>large language models</b> . <sup>13</sup>
<b>Large language model (LLM)</b>	A type of AI system trained on significant amounts of textual data that can generate natural language responses to a wide range of inputs. <sup>14</sup>
<b>Machine learning</b>	A sub-field of <b>artificial intelligence</b> , machine learning is “the process of training a piece of software, called a model, to make useful predictions or generate content from data.” <sup>15</sup>
<b>Model</b>	“[A]ny mathematical construct that processes input <b>data</b> and returns output ... [designed to allow] a system to make predictions.” <sup>16</sup>
<b>Natural Language Processing (NLP)</b>	“A branch of <b>artificial intelligence</b> that helps computers understand, interpret and manipulate human language.” <sup>17</sup>
<b>Non-Generative AI</b>	“AI which primarily relies on <b>algorithms</b> that draw information directly from vast data sets to detect patterns, forecast outcomes and support decisions. The dominant technique is <b>Machine Learning</b> (ML), and its more sophisticated subset, <b>Deep Learning</b> (DL). This kind of AI employs probabilistic models to predict outcomes and give recommendations. [...] The strength of recent non-Generative AI lies in the ability to handle extremely large and potentially unlabelled and unstructured datasets.” <sup>18</sup>
<b>Personal data</b>	Under the UK GDPR, <b>personal data</b> is defined as “any information relating to an identified or identifiable natural person ... such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.” <sup>19</sup>
<b>Probabilistic model</b>	A <b>model</b> that provides multiple possible outputs, based on probability theory, given the same input. <sup>20</sup>

<sup>4</sup> Cambridge Dictionary, *Algorithm* (2024).

<sup>5</sup> IBM, *What is artificial intelligence (AI)* (2024); OpenAI, *Planning for AGI and beyond* (2023); Elliot Jones, *What is a foundation model?* (2023); Google, *Machine Learning Glossary: Artificial general intelligence* (2024).

<sup>6</sup> Regulation (EU) 2016/679 of the European Parliament and of the Council (UK GDPR), art 22(1); ICO, *Rights related to automated decision making including profiling*.

<sup>7</sup> Kate Goddard, Abdul Roudsari and Jeremy C Wyatt, 'Automation bias: a systematic review of frequency, effect mediators, and mitigators' (2012) 19(1) *J Am Med Inform Assoc* 121; Google, *Machine Learning Glossary: Automation bias* (2024).

<sup>8</sup> Cambridge Dictionary, *Data* (2024).

<sup>9</sup> IBM, *What is deep learning?* (2024).

<sup>10</sup> *Machine Learning Models, Decoding Machine Learning Models: Deterministic or Probabilistic?* (2024).

<sup>11</sup> Elliot Jones, *What is a foundation model?* (2023).

<sup>12</sup> Google, *Machine Learning Glossary: Generative AI* (2024).

<sup>13</sup> Jean Kaddour et al, 'Challenges and Applications of Large Language Models' (2023) *arXiv:2307.10169*, 19-20.

<sup>14</sup> Elliot Jones, *What is a foundation model?* (2023).

<sup>15</sup> Google, *What is Machine Learning?* (2024).

<sup>16</sup> Google, *Machine Learning Glossary: Model* (2024); Batta Mahesh, 'Machine Learning Algorithms—A Review' (2019) 9(1) *International Journal of Science and Research* 381.

<sup>17</sup> SAS, *Natural Language Processing (NLP)* (2024).

<sup>18</sup> OECD, *The Impact of Artificial Intelligence on Productivity, Distribution and Growth* (2024), 10.

<sup>19</sup> UK GDPR, art 4(1); ICO, *What is personal data?* (2024).

<sup>20</sup> Zoubin Ghahramani, 'Probabilistic machine learning and artificial intelligence' (2015) 521 *Nature* 452; *Machine Learning Models, Decoding Machine Learning Models: Deterministic or Probabilistic?* (2024).

## b. Who and what is our “justice system”?

The justice system is a whole ecosystem of people, institutions, and services which advise on, enforce, and administer the law. It can be broadly broken into four parts:

- 1. Administrative justice** concerns how the state treats the people who interact with it. It includes issues such as benefits decision-making, housing provision and immigration decisions, and impacts almost all of us in society who are subject to decision-making by public bodies.
- 2. Criminal justice** concerns the investigation, arrest, prosecution, defence, sentencing, punishment, and rehabilitation of those who are suspected or convicted of criminal offences.
- 3. Civil justice** concerns issues where people, business or other organisations are trying to resolve their disputes with each other or ensure they have their rights respected. It covers a very wide area – from simple damaged goods, housing possession matters or small debt recovery claims to large claims between multi-national companies.
- 4. Family justice** concerns the legal regulation of disputes within families or between members of the family and the state. Issues around child arrangements and finances after separation fall into this category (private family justice), as do cases in which a local authority seeks to take a child into care or place them for adoption (public family justice).

**Actors** involved are incredibly broad:

- Individuals, groups, companies, and organisations can all be parties to a challenge of a public body decision or a civil claim, or suspected or charged with a crime. People who are not a party to, or defendant in, the legal proceedings may also be impacted as victims, witnesses or family or community members.
- Legal professionals – lawyers, arbitrators, and other legal professionals help provide legal advice, settle disputes, and represent people in court — both in a paid and charitable (pro bono) capacity, whilst judges decide cases. Non-lawyers also play critical roles - lay magistrates convict and sentence those charged with a crime and for more serious cases, juries, decide on the defendant's guilt. In tribunals judges sometimes decide cases with non-legal expert panellists.
- Institutional actors also play a role — public authorities are the defendant in the administrative justice system; the police, Crown Prosecution Service, and probation services are involved in the criminal justice system; children's services and social workers are involved in the family justice system.

**Other discrete areas** can sit outside the above four parts but still make up a wider understanding of the justice system, for example inquests conducted by coroners' courts, judge-led inquiries, and specialist courts such as the Court of Protection.



## c. What is “Justice AI”?

The potential uses and benefits of AI within the justice system are widespread, supercharged by the advent of generative AI. Justice systems around the world, and the people, institutions and services which make them up, have already begun to use AI in attempts to improve accessibility of legal processes and laws, particularly for lay people, support decision-making of justice system actors and/or streamline processes – both legal processes and the administrative tasks associated with them.

Over the past decade, police forces in the UK have turned to AI to boost their surveillance capacity, to assess risk and to forecast crime.<sup>21</sup> Lawyers use AI for a range of tasks, including administrative assistance, document review, legal research and drafting.<sup>22</sup> Lawyers, alternative dispute resolution professionals and repeat litigants like insurance companies can make use of litigation prediction AI to strategise when to settle cases.<sup>23</sup>

Meanwhile, people with legal problems are seeing more AI tools appear offering low, or no, cost legal information and advice. In addition, many people are already turning to general purpose consumer facing chat bots like ChatGPT.<sup>24</sup> In the judiciary, at least one UK judge has used ChatGPT to produce a summary of the law, which he included in his judgment,<sup>25</sup> whilst judges in the US routinely use AI risk assessments in their bail and sentencing decision-making.<sup>26</sup>

Below is a visual representation of the types of justice AI in development or which already exist worldwide. It is not exhaustive but rather is illustrative of the diversity of actors using AI and the variety of applications.



<sup>21</sup> Law Society, *Algorithm use in the criminal justice system report* (2019); Law Society, *Mapping algorithms in the justice system* (2019); Lina Dencik et al, *Data Scores as Governance: Investigating uses of citizen scoring in public services* (2018); Fair Trials, *Automating Injustice* (2021).

<sup>22</sup> LawTech, *LawTechUK Ecosystem Tracker* (2024).

<sup>23</sup> LexMachina, *How Lex Machina Legal Analytics Works* (2024); Thomas Reuters, *Westlaw Edge Features* (2024); Predictice, *À propos.* (2024).

<sup>24</sup> E.g., Nick Hilborne, *Australia uses AI to help separating couples split assets* (2020); Sheena Vasani, *'Robot lawyer' company faces \$193,000 fine as part of FTC's AI crackdown* (2024); Contend Legal, *Stand up for your legal rights* (2025).

<sup>25</sup> Bianco Castro and John Hyde, *Solicitor condemns judges for staying silent on 'woeful' reforms* (2023).

<sup>26</sup> UNESCO, *Global toolkit on AI and the rule of law for the judiciary* (2023), 138.



### Improving accessibility

Legal chatbots – lay people

Drafting – lay people

Translation – lay people

### Improving accessibility cross over with streamlining process

Translation – courts

Court FAQs chatbot – courts

### Streamlining process

Document review – lawyers

Drafting – lawyers

Scheduling – courts

Anonymisation/redaction – courts, lawyers

Case management – courts, lawyers

Transcription – court

### Streamlining process cross over with supporting decision-making

Judgment drafting – judges

Triaging – courts, police, lawyers

Evidence review – police

Legal research – lawyers, judges, lay people

Virtual assistants – lawyers, judges, lay people

### Supporting decision-making

Risk prediction – police, probation, judges, public decision-maker

Categorisation – prisons

Litigation prediction – lawyers

Crime forecasting – police, public bodies

Image and audio recognition systems – police

Live biometric recognition systems – police

### Supporting decision-making overlap with improving accessibility

ODR platforms – lay people, lawyers



## d. Who is responsible?

Responsibility for the justice system, and the technology used in it, is spread across various actors and institutions in the UK. Each 'actor' in the system has some responsibility specific to their role in the overall system and the 'supply chain' of enabling tools:

- The **Judicial Office** is led by the head of the judiciary, the Lady Chief Justice, and sits separately from Government. It provides policy and operational support to judicial office holders and houses the Judicial College which trains judges.
- In Government, the **Ministry of Justice**, led by the Lord Chancellor, is responsible for the policy and the funding for much of the justice system, including the courts, prisons and probation services. It works with **Government agencies**, like His Majesty's Courts and Tribunal Service and the Legal Aid Agency, to meet those responsibilities. Many justice responsibilities are devolved to the Northern Ireland Executive and the Scottish Government, and Wales has devolved responsibility for several tribunals.
- The **Home Office**, led by the Home Secretary, is responsible for policing and law enforcement, as well as immigration and wider security issues. However, the **UK's 43 police forces operate independently**, and local police forces hold responsibility for and develop many of their own policies. Several bodies and arrangements play a collaborative and coordinating role, such as the College of Policing and the National Police Chiefs' Council.
- Wider responsibilities which impact the justice system are truly cross-governmental, from the Department for Work and Pensions' policies influencing social security appeals, to social care and health policy determining the practical options open to family judges for the most vulnerable children.
- The responsibilities of the **Department for Science, Innovation and Technology** include creating economic growth through innovation and the digital transformation of public services. How the justice system, and those actors and institutions within it, work as an increasingly digitised system in the 21st century will therefore be directly impacted by this department.

There are also several governance bodies and accountability mechanisms which hold the justice system and its actors to account:

- Judges are held accountable for their decisions through **appeals to higher courts**. Personal accountability is limited to ensure judges are free to make their decisions without fear or favour. Judges are, however, personally responsible for their personal conduct, which can be complained about to the **Judicial Complaints Investigations Office**
- Various bodies **regulate the provision of legal services**, for example, the Bar Standards Board and the Solicitors Regulatory Authority, the Office of the Immigration Services Commissioner, and others. They are overseen by the **Legal Services Board**. Other regulators can also have some relevance to the legal services market, for example the Competition and Markets Authority.
- **Parliament** can debate issues within the justice system and can convene specialist committees on justice policy which hear evidence from experts, such as the Justice Committee. Ultimately, Parliament is Sovereign and can change the law which the justice system administers and enforces.
- The **Information Commissioner's Office** regulates the processing of personal data by many private and public actors across the justice system (excluding people, courts or tribunals acting in a judicial capacity).
- The **Equality and Human Rights Commission** regulates many public and some private actors in the justice system in respect of their equality and human rights duties.

**Wider governance is complex.** Much of it involves governance of public authorities, and therefore ultimately overlaps with the administrative justice system (albeit this can lead to further complexities when public functions are outsourced to private actors). However, several further regulatory bodies, independent reviewers and oversight mechanisms are involved:

- The **Criminal Cases Review Commission** looks into criminal cases when appeal routes have been exhausted, and has the power to resubmit the case to the Court of Appeal.
- Several **independent inspectors** have discrete oversight roles for justice system actors, for example, the Independent Chief Inspector of Borders and Immigration and HM's Chief Inspector of Constabulary and Fire & Rescue Services.
- The **Independent Office of Police Conduct** oversees complaints about police conduct, and Police and Crime Commissioners hold individual police forces to account.
- Several other **independent commissioners** focus on particular issues within the justice system and hold a variety of actors to account: for example, the Children's Commissioner, Victims Commissioners, the Domestic Abuse Commissioner.

- **Ombudsman schemes** provide several discrete complaints mechanisms for (mostly public) decision-making.

Finally, a wider set of actors are involved in holding the justice system to account, outside of formal governance mechanisms. This includes **journalists, academia, civil society, and the general public.**

While this report does not assess the current regulatory landscape, it is worth highlighting the difference between a body's regulatory remit and it possessing the internal capability, skills, policy framework and statutory powers to fulfil that remit. Reports have identified "discord and dysfunction" in the regulation of legal services,<sup>27</sup> an unclear, overlapping, and uncoordinated regulatory landscape in criminal justice,<sup>28</sup> failing regulation of equality and human rights in need of refocus,<sup>29</sup> and persistent calls for 'own initiative' investigatory powers for Ombudsmen to fulfil their remit.<sup>30</sup> Therefore, regulators' functional capabilities will determine the efficacy of their oversight of AI and its use.



<sup>27</sup> House of Commons Justice Committee, *Justice Committee sets out recommendations on the regulation of the legal professions to the Lord Chancellor* (2024).

<sup>28</sup> Ibid.

<sup>29</sup> House of Commons Women and Equalities Committee, *Enforcing the Equality Act: the law and the role of the Equality and Human Rights Commission* (2019).

<sup>30</sup> International Ombudsman Institute, *Ombudsman Peer Review of The Parliamentary and Health Service Ombudsman, UK* (2022), 30.



## 2. Context: Why are we talking about it?



## a. The stakes: our justice system and the rule of law

It is immensely important that we get the use of the AI in the justice system right because of the constitutional significance of the justice system and its potential to deeply impact the lives of many. The justice system plays a vital role in upholding the rule of law and, as part of that, individual rights.

The **rule of law** sits alongside parliamentary sovereignty as a fundamental pillar of our constitution.<sup>31</sup> It is not an idealistic or abstract concept without any real consequences for the general public. On the contrary, its importance can hardly be overstated; it is vital for a stable society, functioning democracy and the financial wellbeing of a country.<sup>32</sup> The Government has recognised a need to promote the rule of law both domestically and internationally. Simply put, the rule of law ensures that:

- the law applies and is accessible equally to everyone;
- no one is above the law; and
- the Government must comply with the law and not exercise power arbitrarily.

**Human rights** are a fundamental part of the rule of law within democracies. As Lord Bingham explained:

*“It is a good start for public authorities to observe the letter of the law, but not enough if the law within a country does not protect what are there regarded as the basic entitlements of a human being.”<sup>33</sup>*

These basic entitlements of a human being are found in many international human rights frameworks to which the UK is a signatory and which, in many cases, helped to draft. This includes the European Convention on Human Rights (“**ECHR**”), which is incorporated in UK domestic law by the Human Rights Act 1998, and with which all UK public authorities — including courts — must comply. However, “basic entitlements of human beings”, i.e., rights, can be found in many countries’ customary and common law traditions and enactments far predating the codification of international human rights law in the twentieth century. In the UK, the Magna Carta — and its guarantee against unlawful detention and punishment without due process — provides one of the earliest examples.



<sup>31</sup> AV Dicey, *Introduction to the Study of the Law of the Constitution* (8th edn, Macmillan and Co Ltd 1915), 198.

<sup>32</sup> JUSTICE, *The State We're In: Addressing Threats & Challenges to the Rule of Law* (2023).

<sup>33</sup> Tom Bingham, *The Rule of Law* (London 2010).





The **justice system** plays a vital role in upholding the rule of law and individuals' human rights:

- Courts, tribunals, and the judiciary have a constitutional function to keep arbitrary state power in check through independent, impartial and competent administration of justice.
- Executive actors in the justice system, including the police, prisons and local authorities, must exercise their powers lawfully and uphold their legal duties, ensure they do not unlawfully interfere with rights, and advance positive obligations, for example to prevent discrimination and to protect the right to life.
- Legal professionals and legal service providers supply the legal help and support required to ensure legal rights and remedies are accessible to all, in accordance with professional ethics including independence and an overriding duty to the court.
- Many actors in wider society play oversight roles, from statutory regulators of particular actors to the media and general public. It is these mechanisms of oversight which establish trust that the rule of law is being observed in the justice system, thereby ensuring legitimacy and stability of the law and its institutions in a functioning democracy.

If the use of AI by justice system actors undermines, rather than upholds, the rule of law and human rights, the justice system's very purpose in society is compromised.

However, inaction may also have material consequences. The rule of law in the UK is in a perilous position, and its guardian — the justice system — suffers from many significant and persistent problems: **unmet legal need and legal aid deserts, rising rates of reported crime, disparities in law enforcement of racialised groups, court backlogs and excessive delays, overrun prisons, and a profession which does not reflect the diversity of the public they serve.**<sup>34</sup> It is not a system which was designed as a system but rather has emerged over hundreds of years. It still contains countless blind spots, persistent inequalities and numerous inefficiencies. All these problems require continued attention and reform efforts, without which its role in upholding the rule of law and human rights is threatened and undermined.

As a result, there are many opportunities for innovators, researchers and policy makers to ask how we might use AI to help tackle old and new problems in the justice system. This does not mean that AI is the answer to these problems, but it is likely to have a growing place in the range of tools considered when innovative thinking takes place on key issues, e.g., improving the efficiency of resource-strapped services, supporting the accessibility of legal help, or improving the performance of decision-makers.

---

<sup>34</sup> JUSTICE, *The State We're In: Addressing Threats & Challenges to the Rule of Law* (2023).



## b. The innovation context

AI has been widely heralded as a transformational technology. Its ability to perform complex analytical tasks, and its enhanced potential for autonomy and adaptiveness, distinguish it from other technologies as a step-change which will fundamentally impact the way we live.<sup>35</sup>

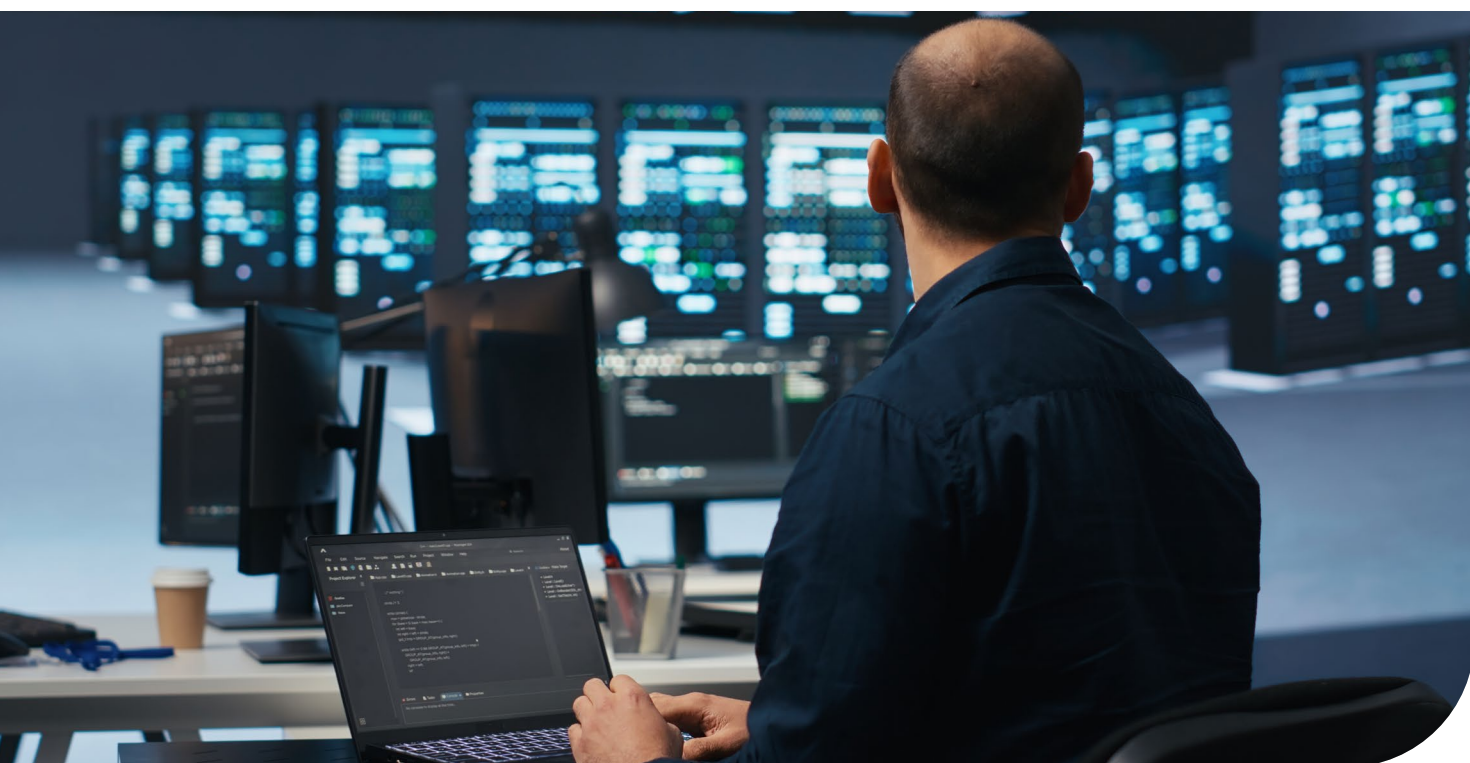
The promise of these features is in their potential to innovate autonomously and to generate productivity. Early evidence supports the expectation that AI could produce productivity growth and lead to higher quality work.<sup>36</sup> As a result, there are active policy discussions and research initiatives into the possibilities for AI across many sectors, from advancements in drug discovery enabled by AI predictions of protein structures,<sup>37</sup> to better climate change planning through seasonal AI forecasting of ice in the Arctic.<sup>38</sup>

The UK Government has clearly stated its aim to seize the opportunities AI presents, “to improve people’s lives, drive productivity and deliver growth”.<sup>39</sup> It is also clear that the justice system is in scope; the Lord Chancellor and Secretary of State for Justice the Rt Hon Shabana Mahmood MP stating:

*“I am very interested in the role that AI and other technology and digital solutions can play in increasing efficiency and productivity in the Crown Court system and the court system more broadly.”<sup>40</sup>*

A further important context is the reliance on, and therefore the power which lies in, the private sector when innovating with AI.

For as long as independent legal services have been available for purchase, the private sector has been part of how justice is delivered. In the current landscape, many fundamental areas of the justice system rely on, or are delivered by, the private sector, including legal services, many alternative dispute resolution options, some prisons and immigration detention centres, court security services, and the digital resources used across the court system, including the common platform and the video hearings service.



<sup>35</sup> Francesco Filippucci et al, [The impact of Artificial Intelligence on productivity, distribution and growth: Key mechanisms, initial evidence and policy challenges](#) (2024), 7 (citing Lipsey, Carlaw and Bekar, [Economic Transformations: General Purpose Technologies and Economic Growth](#) (OUP 2005) who called it the next “general purpose” technology, i.e. technologies which are ubiquitous across society and have a range of technical and economic applications. Previous “general purpose” technologies included computers, the internet, the steam engine, and electricity).

<sup>36</sup> Francesco Filippucci et al (2024) [ibid.](#); Nestor Maslej et al, [Artificial Intelligence Index Report 2024](#) (2024).

<sup>37</sup> John Jumper et al, [Highly accurate protein structure prediction with AlphaFold](#) (2021) 596 *Nature* 583; Josh Abramson et al, [Accurate structure prediction of biomolecular interactions with AlphaFold 3](#) (2024) 630 *Nature* 493.

<sup>38</sup> Tom Andersson et al, [Seasonal Arctic sea ice forecasting with probabilistic deep learning](#) (2021) 12 *Nature Communications* 5124; IceNet, [AI-powered sea ice forecasts](#) (2024).

<sup>39</sup> Department for Science, Innovation and Technology et al, [AI expert to lead Action Plan to ensure UK reaps the benefits of Artificial Intelligence](#) (2024).

<sup>40</sup> Sophie Huskisson, [Artificial intelligence may be used in courts to stop victims waiting years for justice](#) (2024).

Governance mechanisms vary, but there are many ways to secure standards, including ethical codes, regulatory bodies, statutory duties or restrictions, and contract provisions in outsourcing relationships. They are important to get right but, as JUSTICE has found, there are risks. In the context of Government outsourcing, JUSTICE has found several examples of standards slipping, incurring significant risks to individuals' rights and expense to the taxpayer.<sup>41</sup> Too often there is a "hands off" delegation, when what is needed to ensure effective risk management to individuals' rights is collaboration, transparency and accountability.

In the context of AI, the governance challenge is even more complex due to the fact that those who develop the AI tools are often large international companies – the a-symmetry of information, power and capability is significantly tipped in favour of the private entity and not the public administrators. As the Competition and Markets Authority has observed:

***“In AI, we have a disruptive technology that – perhaps for the first time in the history of innovation – is not disrupting the major incumbents who already hold strong market power in some of today’s most important markets, but instead could end up reinforcing their market power. Aggravating the pre-existing tendency towards concentration.”<sup>42</sup>***

The opportunity to use AI to innovate – with the general public and the taxpayer in mind as the ultimate beneficiaries – needs consideration of the risks of doing so while relying on particularly powerful private sector actors, and the available levers and governance mechanisms in place. Reliance on the private sector is not wrong per se, and indeed the public sector meeting all its responsibilities in house is unrealistic. However, it is important to understand that such actors may not align with the duty-driven motives of the public sector. For most private sector companies, owner or shareholder profits are the primary motive, and this will place their motivations at odds with the public good at times. Consequently, governance which is able to mitigate the differences in these drivers is essential.



<sup>41</sup> JUSTICE, *Beyond the Blame Game* (2024).

<sup>42</sup> Marcus Bokkerink, *Speech to the Regulation Forum Chairs' Summit* (2024).

### c. Our approach

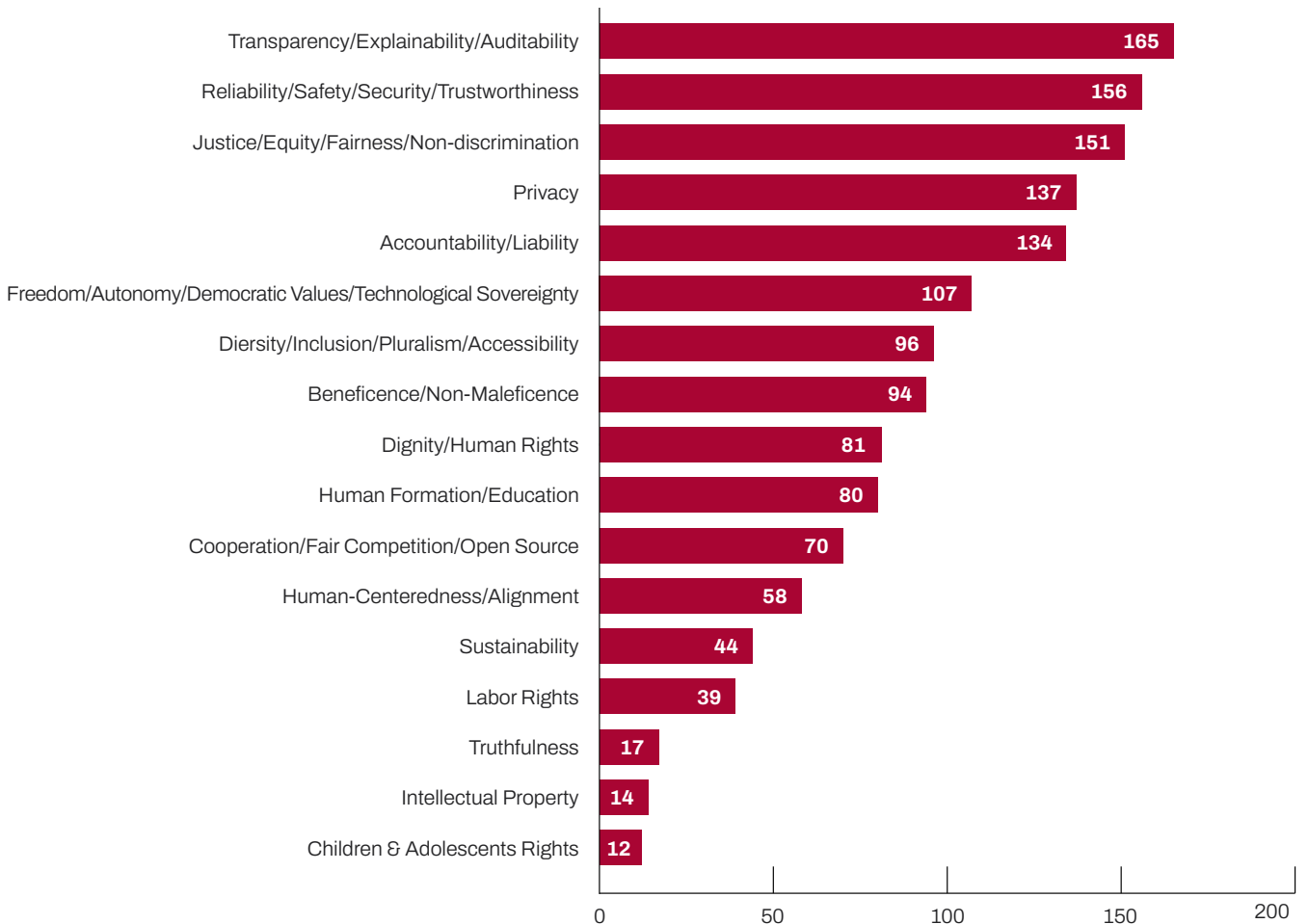
We set out below a framework to guide individuals considering the deployment of AI within the justice system. In compiling this framework, we have taken a rule of law and human rights-based approach that is people centred, but ensures those looking to deploy AI in the justice system take a systemic view of its impacts as well.

We have adopted a rule of law and human rights-based approach because, as articulated above, if the use of AI by justice system actors undermines rather than upholds either or both, the justice system’s very purpose in society is compromised.

Yet, the majority of global conversations about AI have not been framed in terms of human rights and the rule of law. Rather, the swell of activity over the past decade in response to AI has been in the field of AI ethics. The most prevalent global AI ethics are broadly consistent with the most commonly referenced principles documents in the UK, including the Alan Turing Institute’s “FAST” principles (Fairness, Accountability, Safety, Transparency)<sup>43</sup> and the five cross-sectoral principles set out in the UK Government’s 2023 AI White Paper:

- Safety, security and robustness;
- Appropriate transparency and explainability;
- Fairness;
- Accountability and governance;
- Contestability and redress.

**Prevalence of types of principles in international principles statements or guidance documents, Nicholas Kluge Corrêa et al, ‘Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance’ (2023) 4(10) Patterns 100857**





However, beneath such broad reliance on AI ethics is a wide variety of possible approaches, which taken together are incredibly hard to navigate and make sense of within the context of justice. In JUSTICE’s review of several AI ethics principles statements, we identified a breath of different interpretations of common principles;

- **“Fairness”** is usually linked to notions of discrimination or bias, but can range from preventing unlawful discrimination,<sup>44</sup> preventing undesirable, but not necessarily unlawful discrimination,<sup>45</sup> preventing unfair market outcomes,<sup>46</sup> preventing the manipulation of human decision-makers to take decisions they would not have otherwise,<sup>47</sup> promoting equitable outcomes, such as the equitable distribution of advertisements,<sup>48</sup> collecting demographic data, such as age, gender, and race, in order to assess potential discriminatory impact of AI systems,<sup>49</sup> to simply “considering impacts on different groups of stakeholders”.<sup>50</sup>
- **“Safety and security”** can refer to a wide range of considerations, including the need to ensure AI systems operate as expected even with “unexpected changes, anomalies, and perturbations”,<sup>51</sup> ensuring that AI systems prioritise “human life, health, property, and the environment” and align with ESG principles,<sup>52</sup> making AI systems more transparent to allow for a collaborative approach to cybersecurity, for example the promotion of “bug bounties”,<sup>53</sup> or making AI systems less transparent to prevent potential abuse by adversarial actors and slowing down the output of scientific publications.<sup>54</sup>

- **“Transparency and explainability”** can refer to making an AI system “interpretable,” allowing a user to understand why it has made a specific decision,<sup>55</sup> justifying how an AI system is designed,<sup>56</sup> justifying the outcomes given by an AI system,<sup>57</sup> highlighting the limitations of an AI system,<sup>58</sup> examining an AI system’s prediction accuracy, such as through the use of “Local Interpretable Model-Agnostic Explanations” (LIME),<sup>59</sup> increasing output on scientific papers,<sup>60</sup> continuously educating AI users on the systems they are using,<sup>61</sup> producing documentation understandable to both experts and non-experts,<sup>62</sup> or mandating reporting standards.
- **“Accountability and governance”** can refer to creating and following industry-led best practices,<sup>63</sup> creating an industry-led consortium to collaborate with policymakers to create “evidence-based policy recommendations”,<sup>64</sup> giving human justifications for AI decision-making,<sup>65</sup> allowing human review of the creation and implementation of AI systems,<sup>66</sup> taking legal accountability for when things go wrong,<sup>67</sup> indemnifying users against copyright lawsuits for using a specific AI system,<sup>68</sup> taking ethical or moral responsibility for when things go wrong,<sup>69</sup> or reporting the negative impacts from AI systems.<sup>70</sup>

<sup>44</sup> Department for Science, Innovation and Technology and Office for Artificial Intelligence, *A pro-innovation approach to AI regulation* (2023), para 52.

<sup>45</sup> Google, *AI Principles Progress Update* (2023), 3.

<sup>46</sup> Department for Science, Innovation and Technology and Office for Artificial Intelligence, *A pro-innovation approach to AI regulation* (2023), para 52.

<sup>47</sup> Regulation (EU) 2024/1689 of 13 June 2024 laying harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (*Artificial Intelligence Act*) [2024] OJ L, 2024/1689, art 5.

<sup>48</sup> Meta, *Responsible AI* (2024).

<sup>49</sup> Ibid.

<sup>50</sup> AWS, *Responsible AI* (2024).

<sup>51</sup> David Leslie, *Understanding artificial intelligence ethics and safety* (2019), 30.

<sup>52</sup> McKinsey, *Responsible AI (RAI) Principles* (2024).

<sup>53</sup> Google, *Acting on our commitment to safe and secure AI* (2023).

<sup>54</sup> OpenAI, *OpenAI Charter* (2024).

<sup>55</sup> David Leslie, *Understanding artificial intelligence ethics and safety* (2019), 39-40.

<sup>56</sup> Ibid., 24, 35.

<sup>57</sup> Ibid., 24, 36.

<sup>58</sup> OECD, *AI principles* (2024).

<sup>59</sup> IBM, *What is responsible AI?* (2024).

<sup>60</sup> Google, *AI Principles Progress Update* (2023); OpenAI, *OpenAI Charter* (2024).

<sup>61</sup> IBM, *What is responsible AI?* (2024).

<sup>62</sup> Meta, *Responsible AI* (2024).

<sup>63</sup> Ibid.; Google, *AI Principles Progress Update* (2023); AWS, *Responsible AI* (2024).

<sup>64</sup> Meta, *Responsible AI* (2024).

<sup>65</sup> David Leslie, *Understanding artificial intelligence ethics and safety* (2019), 24, 35-36.

<sup>66</sup> Ibid., 35-36.

<sup>67</sup> UNESCO, *Recommendations on the Ethics of Artificial Intelligence* (2022), 23.

<sup>68</sup> Google, *AI Principles Progress Update* (2023).

<sup>69</sup> UNESCO, *Recommendations on the Ethics of Artificial Intelligence* (2022), 23.

<sup>70</sup> European Commission, *Ethics Guidelines for AI* (2019), 20.

Such breadth in ethical considerations has led to criticisms of the “vagueness and elasticity” of the scope, content and impact of these standards,<sup>71</sup> especially when they are interpreted by private actors.

Conversely, there are real benefits in using international human rights standards as the guiding normative framework, rather than ethics, noting human rights standards meet many of the concerns at which AI ethics are aimed.<sup>72</sup>

Human rights are existing norms universally recognised amongst democracies. They have long histories of societal engagement and deliberation in our state institutions. This includes adjudication in the courts, as human rights standards are enforceable to an extent that ethics are not.

The same can be said for additional principles inherent to the rule of law, such as the prevention of arbitrary power and the separation of powers between state institutions. These are protected by public law principles including the requirement for sufficient enquiry, the prohibition against fettered decision-making, the requirement for reasonableness and, when fundamental rights are engaged, proportionality,<sup>73</sup> as well as other principles developed in common law such as open justice.

This rich history means these rights and common law principles are already embedded as cultural and legal norms, they can serve as a common taxonomy of standards, and they are less prone to “rebranding” by different actors.

Further, taking a human rights and rule of law based approach places the focus on the needs of individual rights holders and centres the need to ensure access to justice for all. It therefore promotes a ‘**people centred approach**’, which places the needs of people from the justice system at the heart of the design, development and deployment of AI within it.<sup>75</sup>

The UK has already taken steps to recognise the importance of upholding the rule of law and human rights in AI on the international stage, by ratifying the Council of Europe’s Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law.<sup>76</sup> If human rights and rule of law are important standards for AI across society, then they must be understood as critical in the justice system context.

Although it has its many parts, it is important to see the justice system as a whole and consider the impacts of changes through a systems wide lens.

By taking a **system-wide** approach, it is possible to take a step back from individual use cases and actors to consider the impact of any specific AI development on the justice system, and importantly the wider impacts on the rule of law and human rights. Our review revealed significant AI research and policy examining AI applications in specialised areas such as policing, legal services, and judicial proceedings, but identified a notable gap in comprehensive, system-wide analysis. It is both true that the individual uses in one area of the justice system are likely to have repercussions elsewhere and there is a responsibility on those involved to understand these impacts. Furthermore, the system wide accumulative effect of AI used across the system is likely to change the very nature of the justice system and its interplay in society more broadly – again we argue that there is a responsibility on the part of relevant actors to consider and understand these issues, before the event.

<sup>71</sup> Karen Yeung, Andrew Howes, and Ganna Pogrebna, ‘AI Governance by Human Rights-Centred Design, Deliberation, and Oversight: AN End of Ethics Washing’ in Markus Dubber, Frank Pasquale, and Sunit Das (eds), *The Oxford Handbook of Ethics of AI* (OUP 2020).

<sup>72</sup> *Ibid.*; See also UNHCR, *Artificial intelligence must be grounded in human rights, says High Commissioner* (2023); Mark Latonero, *Governing Artificial Intelligence: Upholding Human Rights & Dignity* (2018); Amnesty International and Access Now, *The Toronto Declaration* (2018); Marc-Antoine Dilhac et al, *The Montréal Declaration for a Responsible Development of Artificial Intelligence* (2017).

<sup>73</sup> Jennifer Cobbe, ‘Administrative law and the machines of government: judicial review of automated public-sector decision-making’ (2019) 39(4) *Legal Studies* 636.

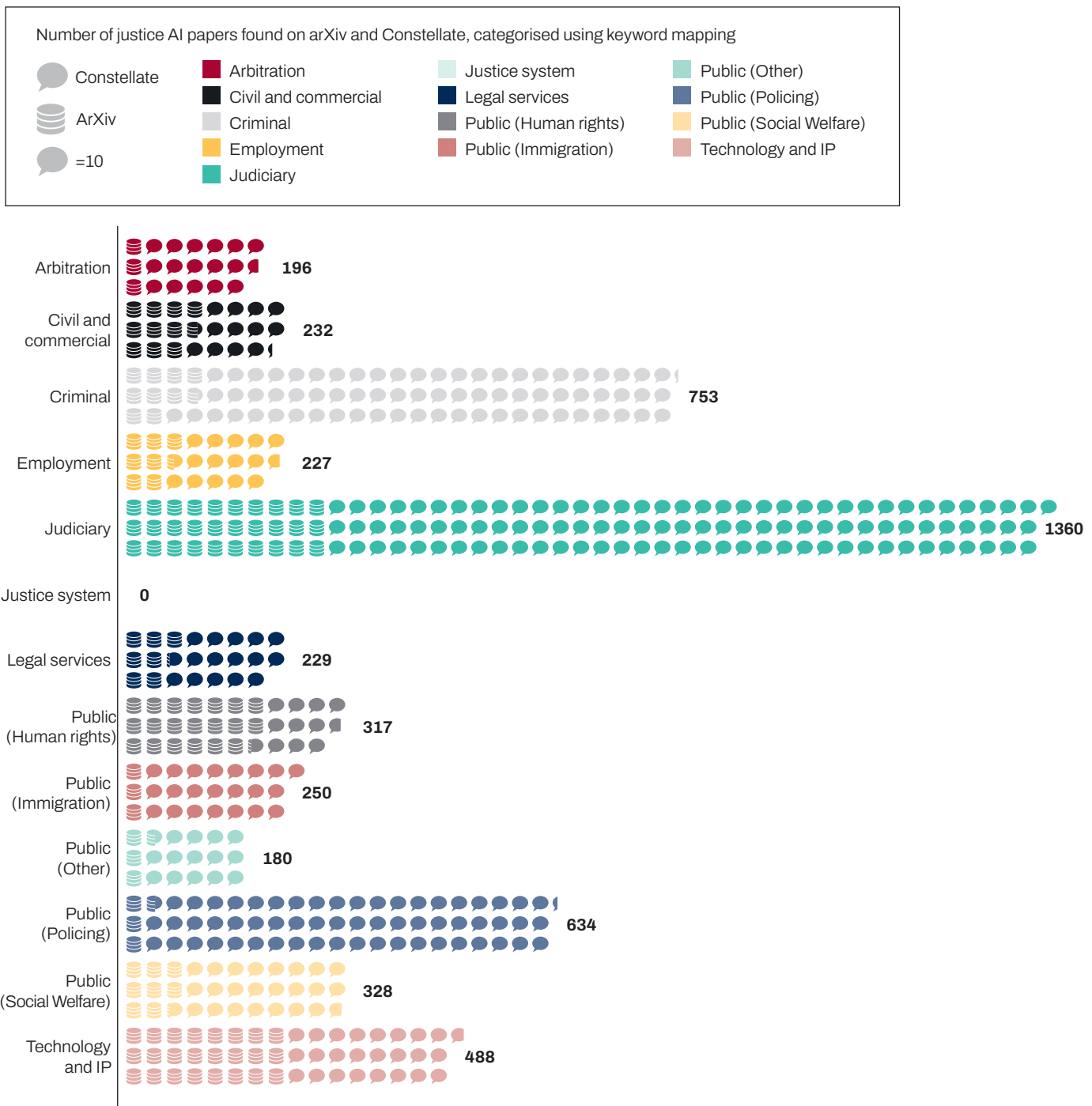
<sup>74</sup> A wider concept in common law than that found in human rights articles, e.g. in the limited reference to public hearings in the right to a fair trial in Article 6 ECHR.

<sup>75</sup> *The OECD Framework and Good Practice Principles for People-Centred Justice* (2021), provides that delivering people-centred justice requires identification of legal needs; state delivery of inclusive, targeted, and appropriate services; a coherent and coordinated response by the justice system; empowerment of citizens to participate in the legal system, such as raising legal awareness; and the establishment of accountability and monitoring mechanisms.

<sup>76</sup> *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (adopted 5 September 2024) 225 Council of Europe Treaty Series 1.

## Many research papers on justice AI focus on particular discrete areas of law

We could not find a single paper in arXiv and Constellate, that reviewed justice system use of AI holistically



Source: arXiv API, Constellate Dataset Builder • ArXiv is an open-access archive of scientific papers operated by Cornell University, whose materials are not peer-reviewed. Constellate datasets include research articles from multiple providers, namely JSTOR, Portico, Chronicling America, Doc South, South Asian Open Archive, and Reveal Digital. Methodology: We searched arXiv and Constellate for all papers on "justice AI", "legal AI", and "law AI" and categorised them by searching for keywords in either their title, executive summary, or tags (if applicable).

In light of the above, we propose that the overarching purpose of any innovation in the justice system must be to **better deliver a justice system which people can trust to uphold the rule of law and protect human rights for everyone**. The next section analyses the opportunities and risks to achieving this with AI.



# 3. Analysis: Goals and opportunities for the use of AI in our justice system

If the purpose of innovation in the justice system is the **better delivery of a system people can trust to uphold the rule of law and protect human rights of everyone**, then justice system actors need to ask whether this purpose is furthered by AI in any given context.

To help justice system actors answer this question, we have identified 3 fundamental goals of the justice system:

1. Equal and effective access to justice;
2. Fair and lawful decision-making; and
3. Scrutiny and public engagement.

If AI can improve the achievability and function of these goals, it can deliver a stronger justice system. **However, if it does not improve these elements, then the justice system is less able to uphold the rule of law, protect human rights, and command the trust of society in doing so.**

In this section, we set out:

- how each goal upholds human rights and the rule of law; and
- the opportunities for AI to further each goal and thereby support a stronger justice system.

In Chapter 4, we set out the risks posed across all three goals and therefore to the justice system as a whole.

# GOAL ONE: Equal and effective access to justice

**In a nutshell:** The justice system is there for all who need it.

## ***i. How this goal upholds human rights and the rule of law***

Access to justice is fundamental to the rule of law. In essence, it requires that individuals are able to access and vindicate their rights or defend themselves against claims or charges. This principle has deep roots dating back to the Magna Carta, which declared that “to no one will we sell, to no one deny or delay right or justice.”

In international human rights law, access to justice is part of the right to a fair trial, found in Article 6 of the ECHR and also protects the right to an effective remedy in Article 13 ECHR.

It is a gateway to the protection of other rights through legal processes: it practically allows people to challenge discrimination, assert their property rights, or access the criminal law’s protection when they have been a victim to a crime, for example, access to justice for domestic abuse victims protects them from inhuman and degrading treatment.

Access to justice is inherently practical: it is about how people can have access to their legal rights in reality, rather than only theoretically. It therefore has many practical aspects, reflecting the barriers which can arise, including:

- the legal and practical information and advice people may require to understand and enforce their legal rights and obligations;
- available formal mechanisms to assert those rights, from reporting crime to making civil claims, and challenging public body decision-making;
- informed and consensual opportunities for settlement, where appropriate;
- an outcome in a reasonable time; and
- decisions are made fairly (which we separately consider under goal two).

## ***ii. Opportunities***

Currently there is rank inequality between those who can and cannot access justice in our justice system. For many, rights and remedies are theoretical and illusive, rather than practical and effective.

The reasons are numerous, from intimidating processes, to cultural barriers, poor understanding of rights and lack of funding. The implementation of the Legal Aid, Sentencing and Punishment of Offenders Act 2012 had a sweeping impact, significantly reducing state assistance for those who cannot afford lawyers.<sup>77</sup>

Moreover, persistent delays and resource pressures remain, as various parts of the system struggle to meet demand.<sup>78</sup> The pressure is not just in the courts: legal aid lawyers repeatedly stress the unsustainability of the legal aid sector in the face of resource pressures. The number of legal aid practitioners available is in decline and so-called “legal aid deserts” mean individuals can be eligible for free help but find themselves unable to access it. Furthermore, police resources to investigate crime are under strain. There is therefore an imperative to act to address inequality and ineffectiveness in access to justice and an understandable appetite to consider how AI may help improve the status quo.

There are already indications that AI may offer new potential to advancing this goal. These activities can be usefully divided into two categories: those which help individuals’ access information or improve their experience of the justice system, and those which increase the capacity of the justice system to provide better and more timely access to justice.

<sup>77</sup> EHRC, *The impact of LASPO on the routes to justice* (2018).

<sup>78</sup> John Hyde, *Employment tribunal hearings delayed to 2026* (2024); Andrew McFarlane, *A View from The President’s Chambers: July 2024* (2024).

## Access to justice may be improved using AI in many ways, for example through:

**1. Legal information:** Accessing legal information and knowledge historically has been a 'high barrier to entry' hence the reliance on experts such as highly trained lawyers. AI systems can provide a wide and growing variety of assistance which lowers this threshold by for example, drawing out insights into legal information contained in judgments,<sup>79</sup> summarising information to help individuals understand their rights and options, and tailoring information presented to particular queries. These tools can be helpful for both individuals and to practitioners.

**2. Procedural support:** AI tools have the ability to help in use cases where there are currently real needs such as assisting litigants in person with tools designed to help them to navigate legal processes,<sup>80</sup> including automated drafting of legal documents and ensuring they meet formal court requirements.<sup>81</sup> AI tools may also show promise in guiding users through the procedural steps, helping users understand and fulfil necessary legal criteria.<sup>82</sup> There is also real potential in delivering better dispute resolution systems across civil and commercial courts, building on the platforms of digitalisation which are a prerequisite to being able to successfully use AI.<sup>83</sup>

**3. Translation technology:** Literacy in the language used by the courts is essential, but for some individuals this is a real barrier. Automated translation technology can assist those who have a different first language to both access and more effectively participate in the process. In business-to-business settings the ability to take documents in one language and have them translated seamlessly has clear advantages in terms of both accuracy and efficiency.

In addition, there is also the potential to improve the timeliness of justice processes, for example by:

**4. Streamlining administrative tasks:** AI has the potential to significantly reduce delays by streamlining various procedures. In courts and tribunals, AI tools can assist in scheduling hearings, managing case files, and automating routine tasks, thereby freeing up judicial resources for more complex matters.<sup>84</sup> Similarly, AI tools can increase the capacity of legal professionals by managing case files and automating routine tasks, thereby increasing their capacity to advise clients.

**5. Document review and due diligence:** AI systems can offer significant efficiencies in document review tasks. For example, technology assisted review systems learn what to look for based on a human expert's review of a small set of documents, then extrapolate those judgments to the remaining document collection. While mostly used in civil litigation,<sup>85</sup> the applications are wider and include criminal document review by the CPS.<sup>86</sup>

<sup>79</sup> Václav Janeček, 'Judgments as bulk data' (2023) 10(1) *Big Data & Society*.

<sup>80</sup> Hannes Westerman, 'Using Artificial Intelligence to Increase Access to Justice' (LLD thesis, University of Montreal 2023); Hannes Westerman and Karim Benyekhlef, 'JusticeBot: A Methodology for Building Augmented Intelligence Tools for Laypeople to Increase Access to Justice' (2023) arXiv:2308.02032.

<sup>81</sup> Hannes Westerman, 'Using Artificial Intelligence to Increase Access to Justice' (LLD thesis, University of Montreal 2023).

<sup>82</sup> Gráinne McKeever, John McCord and Mark Potkewitz, *Understanding and supporting legal participation for Litigants in Person* (2019); Felix Steffek et al, *Access to justice through artificial intelligence* (2023).

<sup>83</sup> Michael Cross, *Canada shows way to online claims future* (2018); Geoffrey Vos, *Speech by the Master of the Rolls: The Future of Courts* (2024).

<sup>84</sup> Court of Justice of the European Union, *Artificial Intelligence Strategy* (2023).

<sup>85</sup> *Pyrrho Investments Ltd v MWB Property Ltd* [2016] EWHC 256 (Ch).

<sup>86</sup> NPCC, College of Policing and CPS, *National Disclosure Improvement Plan* (2020).



## GOAL TWO: Fair and lawful decision-making

**In a nutshell:** Decisions are made fairly and in accordance with legal limits.

### ***i. How this goal upholds human rights and the rule of law***

Several actors wield substantial powers in the justice system which shape people's lives. Police arrest and investigate; lawyers provide legal advice on which people rely; judges, magistrates and tribunals order financial awards, sever family ties, and imprison; and juries collectively determine guilt. Wherever these powers lie, people within the justice system are making decisions which will have a direct impact on the lives of individuals. These decisions must therefore be fairly and lawfully made.

Independence, impartiality and competence make up "fair" judicial decision-making and are central to many international human rights treaties' descriptions of a "fair trial".<sup>87</sup> Meanwhile the independence, impartiality and competence of the legal profession is equally fundamental to the preservation, promotion and safeguarding of the rule of law.<sup>88</sup>

These qualities of decision-making are also relevant to state actors. Competence of professionals is of course unequivocally important throughout the system.

In the context of policing, operational independence is a "fundamental principle",<sup>89</sup> while the Peelian Principles recognise the importance of "absolute impartial service to the law".<sup>90</sup> In other public office roles in the justice system, from probation officers to Home Office Presentation Officers in the immigration and asylum tribunal, are underpinned by the Nolan Principles of Public Life, which include impartiality and guarding against inappropriate influence.

Furthermore, there is a clear link, between the "impartiality" of decision-making across the justice system and preventing discrimination. Equality before the law and non-discrimination are key tenets of the rule of law, and discrimination is prohibited for the enjoyment of human rights in Article 13 ECHR, and by the public and private sectors under the Equality Act 2010.

### **Decision-making must also be lawful.**

Decision-makers – including the judiciary and public sector decision-makers – must not exercise power beyond their legal scope, do so arbitrarily, or fetter their discretion. In all such circumstances, fundamental human rights must be protected. This means no powers are exercised in breach of absolute rights or guarantees, and when rights are not absolute and can be limited, this is only done in accordance with the law, when necessary for a legitimate aim, and proportionately.

If decision-making is not fair and lawful, there can be no meaningful protection of the rule of law or human rights in our justice system. Such decision-making puts individuals at risk of injustice while undermining the legitimacy of various actors. Given the powers at the disposal of many justice system actors, the human rights consequences of this would be severe.



<sup>87</sup> Universal Declaration on Human Rights (adopted 10 December 1948) UNGA Res 217 A(III) (UDHR), art 10; International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) UNTS 171 (ICCPR), art 14; Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR), art 6.

<sup>88</sup> Commonwealth Bar Leaders, [Declaration on preserving and strengthening the independent of the judiciary and on ensuring the independence of the legal profession](#) (2023).

<sup>89</sup> *R v Commissioner of Police of the Metropolis, Ex parte Blackburn* [1968] 2 QB 118, CA.

<sup>90</sup> Home Office, [Definition of policing by consent](#) (2012), principle 5.

## ii. Opportunities

Resource pressures and backlogs across the justice system are putting strain on the system as a whole and individual professionals' capacity. AI has the potential to support decision-makers, allowing them to have at their disposal more data-derived insights than they would otherwise have, reducing their procedural workload, or by presenting research summaries to supplement and augment their knowledge. Because AI can process data on a scale that humans cannot, AI may be able to uncover patterns and insights that humans would not otherwise be able to.<sup>91</sup>

### Examples include:

- 1. Research assistance:** Judges and lawyers can be assisted by AI tools to help them perform tasks, such as legal research or drafting. This can enhance competence in decision-making, either by better informing such decision-making or by increasing the time available for legal analysis.
- 2. Investigative capacity and insights:** police may seek assistance from AI systems to improve their ability to investigate criminal activity. For example, AI has been used to identify victims and perpetrators of child sexual abuse in online images, increasing the scale and speed of analysis while also having the additional benefit of protecting staff welfare by reducing their exposure to distressing content.<sup>92</sup>

There may also be opportunities to help enhance the fairness of decision-makers and therefore improve outcomes for those who experience disparities in the justice system currently. There is significant evidence of unequal treatment and outcomes for Black and minoritised ethnicities in the criminal justice system,<sup>93</sup> and growing research on disparities in other areas such as family justice.<sup>94</sup> While not all discriminatory outcomes are caused by biased decision-making, cognitive biases in decision-making represent a particularly complex challenge, reflecting an interplay of individual prejudices and systemic issues that affect judgment under conditions of uncertainty.<sup>95</sup>

**3. Combating bias:** There have been some efforts globally to use AI-based tools to identify biases: a novel strategy in the US utilised natural language processing to measure the different gender attitudes of judges, by analysing how frequently judges linked men with careers and women with families in their written opinions.<sup>96</sup>

Such AI innovations have the potential to deliver fairer decisions thereby promoting the rule of law and the protection of rights.

<sup>91</sup> For example, in the medical field, AI based algorithms can detect melanoma from dermoscopy images more accurately than dermatologists: Raj Patel et al, 'Analysis of Artificial Intelligence-Based Approaches Applied to Non-Invasive Imaging for Early Detection of Melanoma: A Systematic Review' (2023) 15(19) *Cancers (Basel)* 4694.

<sup>92</sup> Home Office and Sajid Javid, *Pioneering new tools to be rolled out in fight against child abusers* (2019).

<sup>93</sup> JUSTICE, *Tackling Racial Injustice: Children and the Youth Justice System* (2021).

<sup>94</sup> E.g., higher intervention and adoption rates for white children in social care, and higher rates of deprivation of liberty for Black and Asian children. See Charlotte Edney, Bachar Alrouh and Mariam Abouelenin, *Ethnicity of children in care and supervision proceedings in England* (2023).

<sup>95</sup> Amos Tversky and Daniel Kahneman, 'Judgment under Uncertainty: Heuristics and Biases' (1974) 185(4157) *Science* 1124.

<sup>96</sup> Elliott Ash, Daniel Chen and Arianna Ornaghi, 'Gender Attitudes in the Judiciary: Evidence from US Circuit Courts' (2024) 16(1) *American Economic Journal: Applied Economics* 314.

## GOAL THREE: Scrutiny and public engagement

**In a nutshell:** The justice system is open, enabling scrutiny to identify and improve problems and engendering trust and confidence of the public.

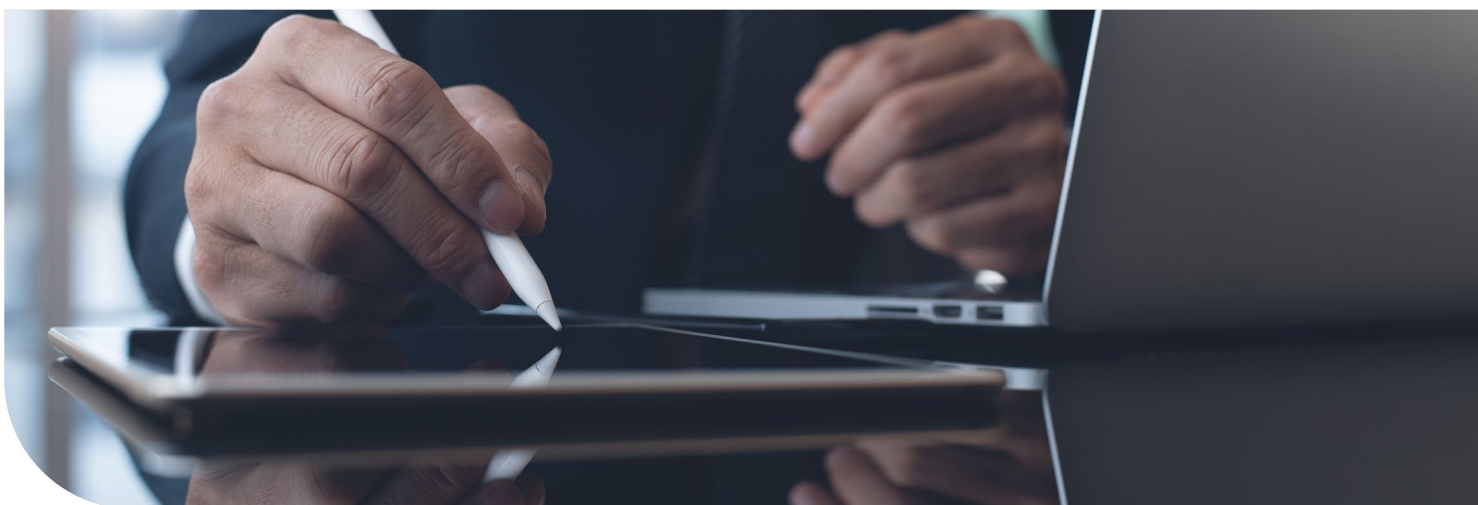
### ***i. How this goal upholds human rights and the rule of law***

Open justice is fundamental to the rule of law. This is reflected in the often-cited phrase “justice should not only be done but should manifestly and undoubtedly be seen to be done.”<sup>97</sup> Open justice is not merely good practice but supports the accountability of justice actors to society at large, on whose behalf they administer the law. As explained by Toulson LJ:

*“How is the rule of law itself to be policed? It is an age-old question. Quis custodiet ipsos custodes - who will guard the guards themselves? In a democracy, where power depends on the consent of the people governed, the answer must lie in the transparency of the legal process.”<sup>98</sup>*

Scrutiny and public engagement are also important to open and transparent government, reflected in the Nolan Principles for Public Life in the principles of honesty, openness and accountability and the Peelian Principles, as mentioned above, which derive legitimacy from openness to, and co-operation with, the public.

Openness therefore has an important role to play in engendering trust and confidence in the justice system and its actors, the legitimacy that lends, and the resultant stability in society.<sup>99</sup> However, scrutiny does not always identify positive aspects of the justice system – it can also identify poor practice, inefficiencies, and unfair outcomes. This is as much the purpose of scrutiny as trust and confidence: “letting in the light and allowing the public to scrutinise the workings of the law, for better or for worse”.<sup>100</sup> Such scrutiny which identified “for worse” would be of little use if nothing were then done to make improvements. Scrutiny of our justice system therefore also has a very practical purpose - to identify need for improvements to deliver a better functioning justice system. Where problems are identified, public engagement will then be required, to understand how change can better serve the public, and how to build back any trust which may have been lost.



<sup>97</sup> Lord Hewart CJ, *R v Sussex Magistrates, Ex parte McCarthy* [1924] 1 KB 256.

<sup>98</sup> Lord Justice Toulson, *R (Guardian News and Media Ltd) v City of Westminster Magistrates' Court* [2012] EWCA Civ 420.

<sup>99</sup> The perception of fairness in criminal justice – both police and courts – has been shown to increase the legitimacy attributed to the institutions involved, and the likelihood of people to accept unfavourable outcomes. See Tom Tyler, 'Procedural Justice, Legitimacy, and the Effective Rule of Law' (2003) 30 *Crime and Justice* 283.

<sup>100</sup> *R (Guardian News and Media Ltd) v City of Westminster Magistrates' Court* [2013] QB 618, para 1.



## ii. Opportunities

In practice, scrutiny and public engagement can come in several shapes and sizes, be it through independent oversight, access to judgments and reasons, academic analysis, court reporters or lay public observers. **There are therefore several opportunities for AI to improve the status quo:**

**1. Publishing judgments:** written and published court judgments contribute to the principle of open justice and access to justice. However, courts are overburdened by the volume of work in the system, which impacts their ability to publish judgments particularly when they may need anonymising.<sup>101</sup> There is some promise in the use of AI to record transcripts of oral judgments, as well as aiding in tasks such as anonymisation and drafting.<sup>102</sup>

**2. Analysis of linked justice data:** In recent years linked administrative datasets from the justice system and other state services have been made available for research through initiatives such as the Ministry of Justice's Data First programme (funded by ADR UK), the Ministry of Justice's Better Outcomes through Linked Data (BOLD) programme, and the Family Justice Data Partnership between Swansea and Lancaster Universities.<sup>103</sup> Facilitating academic research in our justice system helps knowledge of the impacts and outcomes of the justice system, including systemic issues, and helps identify ways to change in the future. The secure platforms – SAIL Databank and the ONS Secure Research Service – and clear data standards in these schemes allow researchers to use computational algorithmic analysis on datasets confidently and responsibly.

**3. Translation technology:** Literacy in the language used by the courts is essential, but for some individuals this is a real barrier. Automated translation technology can assist those who have a different first language to both access and more effectively participate in the process. In business-to-business settings the ability to take documents in one language and have them translated seamlessly has clear advantages in terms of both accuracy and efficiency.

**3. Better facilitation of court observation:** Improving the pace of publication and ease of access to court documents, transcripts and daily lists through AI tools could contribute to the facilitation of public access to the courts, including by the media.<sup>104</sup> Public observation of justice plays a vital role in ensuring scrutiny and public understanding of the justice system, and journalists do so as the “eyes and ears of the public”.<sup>105</sup>

**4. Deliberative democracy tools:** there are examples of innovative uses of AI systems to facilitate better engagement in law-making and policy development, most notably in Taiwan.<sup>106</sup> Such tools could improve the engagement, and trust, of the public in our justice system - in Taiwan increased democratic engagement was correlated with a significant increase in public trust in Government.<sup>107</sup>

**5. Enhancing advocacy for underrepresented and under resourced groups:** AI has been used by advocacy groups to analyse video interviews of incarcerated individuals and present the policy implications which follow, ensuring the voices of those who are usually marginalised in the policy-making process are amplified by groups who otherwise would not have the resources to analyse and produce such materials.<sup>108</sup>

<sup>101</sup> House of Commons Justice Committee, *Open justice: court reporting in the digital age* (2022).

<sup>102</sup> *Ibid.* An example of court transcription, see JUST: Access, *About Us* (2024).

<sup>103</sup> See Ministry of Justice, *Ministry of Justice: Data First* (2020); Ministry of Justice, *Ministry of Justice: Better Outcomes through Linked Data (BOLD)* (2022); Centre for Child & Family Justice Research, *The Family Justice Data Partnership (FJDP)* (2024).

<sup>104</sup> *Ibid.*

<sup>105</sup> House of Commons Justice Committee, *Open justice: court reporting in the digital age* (2022).

<sup>106</sup> Audrey Tang, *Democracy in the age of AI* (2024).

<sup>107</sup> Poly Curtis, *How Taiwan bucked a global trend – and restored voters' trust in politics* (2024).

<sup>108</sup> Brittney Gallagher et al, *Using AI to Give People a Voice, a Case Study in Michigan* (2023).

# 4. Managing risks: Innovating with AI to support the justice system

People need a justice system that works, and one they can trust to uphold the rule of law and protect human rights. As outlined in the above chapter, there is significant potential to help meet this need through the deployment of AI, but success is far from guaranteed. There are many well documented risks which need to be considered and assessed in the context of the justice system. Some of these risks can be mitigated entirely, others to a degree and in some cases not in any meaningful way.

To ensure innovation makes the most of the opportunities while guarding against risks, the rule of law and human rights cannot be afterthoughts or distant considerations: they must be embedded, both practically and in policy terms, in the innovation approach adopted.

To achieve this successfully we propose two requirements which act as safeguards and prevent unintended harms.

- 1. Genuine Improvement:** AI systems and tools developed for use in the justice system should have a clear objective of improving one or more of the core fundamental goals of a well-functioning justice system, as set out in Chapter 3.
- 2. Duty to act responsibly:** All those involved in the design, development and deployment of AI within the justice system have a responsibility to ensure that the core features of the rule of law and human rights are embedded in each stage.

## a. Goal-led

Is the proposed use of AI advancing one or more of the goals of the justice system, by either solving an existing problem or improving the system in some other way?

This is the first question which should be asked at every stage of ideation, design, development, testing, deployment, upgrade and replacement.

As set out in Chapter 3 we have identified 3 fundamental goals of the justice system:

1. Equal and effective access to justice;
2. Fair and lawful decision-making; and
3. Scrutiny and public engagement.

Each of these goals has hundreds and thousands of use cases, and corresponding existing problems which needs solving, some of which can be assisted, or even solved by the use of AI. Embedding these goals at the start of any AI innovation process will ensure that those developing and deploying AI in the justice system are not focused on change for change's sake, but rather on the purpose of that

change and how the justice system will be improved as a result. This test helps to guard against wasted effort on solving less important or irrelevant use cases or the further embedding of existing problems. 'Kicking the tyres' of an idea to test whether it solves a genuine problem in the justice system and one which relates to one or more of the goals is pro innovation, weeding out weaker ideas, and is supportive of a stronger justice system.

## b. Duty to act responsibly

The justice system is not the place to move fast and break things. The stakes are too high and the risks multi-faceted. All those involved in AI design, development and deployment within the justice system should be obliged to ensure that the rule of law and human rights are embedded at each stage. Careful identification and assessment of risk, and the circumstances that enhance or reduce those risks, should be hardwired into the processes of development and deployment. Failure to manage risks increases the likelihood of AI undermining the justice system. The process of assessment, in order to be robust and credible, must also include the possibility of pausing, redeveloping, rethinking deployment, and in some cases even stopping if significant risks to the rule of law and human rights are identified.





## c. Understanding the Risks

Risks can arise at number of different levels:

- On a **technical level** there are likely to be a number of risks that are inherent within the technology itself related to the quality of the inputs used in the development of the AI systems as well as the technical limitations of the system itself.
- On an **individual level**, the nature and severity of the risk will depend on a variety of factors relating to the characteristics of the individuals who are utilising, or are subject to the AI, as well as the availability of mechanisms to challenge the use of AI.
- Whilst the use of AI may further one of the Goals for a particular group of people, it may also produce inequalities and unintended consequences at a **societal or systemic level**.

Below we illustrate some of the key risks under each of these categories that those seeking to deploy AI in the justice system should consider. Often a crystallised risk is a combination of contributing factors from a number of sources, and as such while the distinctions between these categories of risk are not always clear cut, it is nonetheless helpful in identifying and assessing risks to begin with this categorisation.

Each of these categories of risks, and the specific examples cited below, have the potential to impact any or all of the goals of the justice system. In each case, if the risk is not understood or mitigated, there is a real chance of it undermining access to justice, preventing lawful and fair decision-making or increasing the opacity of the system and eroding confidence and trust. Often the complexities result in multifaceted risks and equally multifaceted impacts. For example, risks associated with a lack of data transparency directly impact on confidence and trust in the system but can also relate to the fairness of decision-making, and in some circumstances even impact on access to justice. Taking a holistic approach to examining potential risks and impacts is essential if the process is to be genuinely effective and credible.

### i. Technical risk factors

The use of AI technologies in the justice system presents a range of technical risks that arise from issues such as the quality of the data used, the reliability of the AI models' outputs, and the clarity and transparency of the processes behind them. Addressing these challenges is essential to ensure that AI tools support — rather than erode — the principles of justice.

#### Poor or incomplete data

High-quality, comprehensive data is the cornerstone of reliable AI-driven insights. However, the justice system has more data gaps than any other public service.<sup>109</sup> The justice system faces significant data challenges, including inconsistent data standards, limited interoperability, and substantial information gaps. More than many other public services, justice-sector data is often fragmented, managed in silos, and not readily available in a standardised format. This is partly due to the fact that responsibility for the justice system is fragmented across a number of Government departments – the Ministry of Justice, Home Office and the Attorney General's Office, as well as arms lengths bodies for example, His Majesty's Prison and Probation Service and His Majesty's Court and Tribunal Service, as well as the independent judiciary and CPS and private actors providing a range of advice, information, support and dispute resolution services. Further, if AI tools are trained on legal information, such as legislation and judicial decisions, the quality and meaning of these data sources must be carefully considered.<sup>110</sup> Legal judgments should not be understood as absolute truth;<sup>111</sup> judges inevitably make decisions under uncertainty with imperfect information (only as presented by the parties and their legal counsel) and may imperfectly use it.<sup>112</sup> Some decisions are later overturned on appeal, while others must be understood within broader legal and factual contexts. Experienced lawyers inherently account for these nuances when using legal corpora, but AI systems may fail to do so unless bulk data is carefully curated and contextualised.

<sup>109</sup> JUSTICE has previously highlighted the impact of such gaps in undermining our knowledge of how our justice system is working JUSTICE, *The State We're In: Addressing Threats & Challenges to the Rule of Law* (2023), 72 onwards; Thomas Pope, Gavin Freeguard, Sophie Metcalfe, *Doing Data justice: Improving how data is collected, managed and used in the justice system* (2023).

<sup>110</sup> Ministry of Justice, HM Courts & Tribunals Service and James Cartlidge, *Court judgments made accessible to all at The National Archives* (2022).

<sup>111</sup> Barry Smith, 'Of Truth and Certainty in the Law: Reflections on the Legal Method' (1985) 30(1) *American Journal of Jurisprudence* 97; Ronald Heiner, 'Imperfect Decisions and the Law: On the Evolution of Legal Precedent and Rules' (1986) 15(2) *Journal of Legal Studies* 227; Jules Coleman, *Truth and Objectivity in Law* (1995) 1 *Legal Theory* 33.

<sup>112</sup> Ronald Heiner (1986), 233 *ibid*; Holli Sargeant and Måns Magnusson, 'Bias in Legal Data for Generative AI' (International Conference on Machine Learning 2024, Vienna, July 2024); Andreas s Östling et al, *The Cambridge Law Corpus: A Dataset for Legal AI Research* (2024) arXiv:2309.12269.

## Algorithmic bias and discrimination

When training data is incomplete, poorly curated, or reflects social inequalities, the resulting models can unintentionally replicate or even amplify these biases.<sup>113</sup> Historical discrimination, skewed enforcement practices, or unequal access to legal resources may be “baked into” the training data, influencing predictions and decisions in ways that disadvantage certain groups. Over time, these biased outputs can create “feedback loops,” where decisions informed by biased data generate new, similarly skewed data. This, in turn, further entrenches unfair outcomes, making it more difficult for marginalised communities to receive equitable treatment and access to justice. Allowing algorithmic bias to remain unchecked can produce and entrench discrimination. It erodes trust in the fairness and legitimacy of AI-driven legal processes and undermines the fundamental principle of equal treatment under the law.

## Uncertain or unpredictable AI methods

Many AI models – including those used for risk assessments, sentencing recommendations, or fraud detection – rely on probabilistic methods. Instead of offering guaranteed correctness, they provide predictions with varying degrees of confidence, which means there is always a margin of error. Decisions informed by these probabilistic systems can have significant real-world implications – including deprivation of liberty, removal of welfare benefits or removal of children from parents’ care. This uncertainty can lead to adverse outcomes if decision-makers, unaware of the underlying probabilities and limitations, treat AI-generated results as fully accurate or certain. Over-reliance on these systems may result in incorrect assessments, misguided enforcement actions, or unjustifiably harsh sentencing decisions.

## Misleading or hallucinated AI information

In addition to issues of uncertainty, some AI models, particularly large language models (LLMs), can produce entirely fabricated or misleading content. Known as “hallucinations,” these outputs may appear coherent and authoritative but are not grounded in reality. This problem can arise in legal research tools driven by LLMs. For instance, recent Stanford research found that tools designed to assist with legal inquiries frequently generated fabricated case citations and inaccurate legal interpretations.<sup>114</sup>

While these errors can appear convincingly authentic, they undermine trust in the technology and can mislead users who rely on AI assistance for interpreting laws, locating relevant jurisprudence, or preparing legal arguments.

## Lack of transparency of models

Many AI models, particularly those based on deep learning, operate as “black boxes,” producing conclusions through complex statistical relationships that are not easily interpreted.<sup>115</sup> This opacity can be further compounded by proprietary technologies and data sources. When an AI system is commercially developed and sold to justice-sector institutions, the underlying algorithms, data processing techniques, and training methodologies may be considered trade secrets. In these cases, vendors provide decision-makers with predictions, assessments, or recommendations without disclosing how those determinations are made. As a result, it becomes technically difficult to pinpoint the factors influencing the model’s outputs or to verify their accuracy and fairness.

This poses particular issues in the justice system context where opacity can limit the ability of decision-makers to understand, or spot errors in, information they are given and exercise their discretion with independence, impartiality and competence. If AI systems limit independence, impartiality and competence, the fairness of decision-making is undermined. This poses two possible avenues for further exploration. First, if accepted as the future norm, it is important in policy terms that there is full consideration of what this means for the concept of a fair and independent decision-making, including issues such as equality of arms and the ability to appropriately defend or respond to a case. Second, if it is not possible to square away the negative impacts in other ways, what thresholds of acceptance of black box systems are applied in the justice sector? Although difficult and in some cases impossible, it is feasible to build AI systems which can offer greater degrees of explanation and transparency, if not in the actual decision-making, then in the process and assumptions used to create the systems. In a public service so vital for the functioning of a society, is it too hard to consider that ‘absolute black boxes’ should not be used? How are the benefits they afford traded off against the goals of the justice system?

<sup>113</sup> See, for example, Joy Buolamwini and Timnit Gebru, ‘Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification’ (2018) 81 Proceedings of Machine Learning Research 77.

<sup>114</sup> Varun Magesh et al, *AI on Trial: Legal Models Hallucinate in 1 out of 6 (or More) Benchmarking Queries* (2024).

<sup>115</sup> Christoph Molnar, Giuseppe Casalicchio and Bernd Bischl, ‘Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges’ (Workshops of the European Conference on Machine Learning 2020, Ghent, February 2021), 417-431; Cynthia Rudin, ‘Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead’ (2019) 1 Nature Machine Intelligence 206.

## ii. Individual harm risk factors

When AI systems are deployed with respect to specific individuals, several factors can limit or exacerbate the risk of harm. Individuals impacted by others' use of AI tools can include those subject to police powers, those whose lawyer has used AI, or litigants in a case in which another actor(s), such as the other party, the court administrators, or indeed the judge are using AI. The circumstances of those individuals will exacerbate or mitigate the risks posed by use of AI and should be considered both at the point of development and deployment.

### Threats to fundamental rights

The deployment of AI technologies in the justice system carries significant risks for fundamental rights. Because justice system actors have at their disposal powers that can profoundly affect individual lives, erroneous AI outputs can directly lead to breaches of these rights. The risk of harm will therefore vary depending on what is at stake for the individual – the potential risk of harm when using AI to help streamline listing for example will be of a completely different order of magnitude to the use of AI in circumstances where fundamental rights are at stake.

In the criminal justice sphere, for instance, the use of inaccurate AI-driven risk assessments may lead to unlawful or arbitrary arrests or detentions, undermining the **presumption of innocence** and breaching the **right to liberty**.<sup>116</sup> Similarly, in the context of immigration and asylum, the use of AI systems which produce biased or inaccurate outputs may result in unlawful or arbitrary detention or even refoulement.<sup>117</sup> In the context of local authority children's services, inaccuracies in AI outputs risks unlawful or arbitrary breach of the **right to a family life**.<sup>118</sup>

Another important threat to fundamental rights arises if an AI tool involves the collection, storage, and processing of personal data and sensitive information, where the **right to privacy** of the data subjects will be engaged. Such tools in the hands of all actors will need to be compliant with data protection law, while responsible public authorities will need to ensure that any interference with individuals' privacy is lawful, necessary in pursuance of a legitimate aim in a democratic society, and be proportionate to that aim. AI models can leverage vast datasets of sensitive information from biometric data to personal, financial and legal records. Such use may infringe on individual privacy rights, especially if the data is collected or processed in ways that are disproportionate or not disclosed to data subjects.

### Inability to reverse detriment

Similarly the severity of risk will vary depending on whether the possible negative impacts of the use of AI on an individual are remediable; if so this represents a lower risk use case. In comparison, cases where the use of the use of AI will result in irreparable harm – such as being charged with an offence or imprisonment – the risks are higher. In these cases, any challenge would not be able to reverse the consequences of the harm, only provide compensation.

### Barriers to fair access to justice

Inaccurate, biased, or poor-quality AI outputs may discourage individuals from asserting their rights or seeking redress. For example, erroneous predictions could deter individuals from bringing a valid claim or challenging an unjust decision, as they may believe the system is stacked against them. This “chilling effect” can undermine the fundamental principle that everyone should be able to vindicate their rights before an impartial tribunal. The ability to ‘spot’ inaccurate or poor-quality outputs is skill dependent and unlikely to be a capability that most individuals, whether lay or expert lawyers, are likely to have.

<sup>116</sup> Sophia Adams-Bhatti and Holli Sargeant, 'Algorithms in the Justice System: Current Practices, Legal and Ethical Challenges' in Matt Hervey and Matthew Lavy (eds), *The Law of Artificial Intelligence* (2nd edn, Sweet & Maxwell 2024).

<sup>117</sup> Madeleine Forster, *Refugee Protection in the Artificial Intelligence Era: A Test Case for Rights* (2022).

<sup>118</sup> Dutch News, *Childcare benefit victims criticise slow pace of redress* (2024).

<sup>119</sup> European Commission for the Efficiency of Justice (CEPEJ), *CEPEJ European Ethical Charter on the Use of Artificial Intelligence (AI) in Judicial Systems and Their Environment* (2018).



## Inequality of arms and digital exclusion

Where one party to litigation or negotiation has access to advanced AI systems and the other does not – because they are less technologically proficient or unable to afford high-cost AI tools – the position of the less powerful party risks being worsened. The European Ethical Charter on AI in judicial systems has warned against such imbalances, emphasising that technological tools should facilitate proceedings without disadvantaging those less familiar with digital means.<sup>119</sup> Although financial means have always played a role in access to justice (for example who can afford lawyers), the orders of magnitude of advantage that AI could present could tip the scales significantly.<sup>120</sup>

## Lack of transparency of AI decisions

A primary issue for individuals affected by AI tools is simply knowing that such a tool has been used in their case. Without that basic knowledge – regardless of an individual’s capabilities and knowledge, they will be unable to understand and scrutinise the decision made and the influence of AI on the decision.

With respect to decisions made by public authorities including law enforcement, the lack of transparency often risks breaching the public law

principle of providing adequate reasons for decisions. In many contexts, this principle is secured by legal requirements (for example, the police must provide reasons for an arrest).<sup>121</sup>

In a judicial decision-making setting, a lack of transparency would undermine a party's right to a fair trial, including the right to a reasoned judgment and the ability to appeal. While some argue that judicial reasoning can also be opaque, human decision-makers can be questioned, held accountable, and rebuked for discriminatory behaviour, ensuring a level of scrutiny that is difficult to apply to AI systems.<sup>122</sup>

Although there has been some recognition of the importance of transparency in this context progress is slow, and without clear regulatory expectations and associated enforcement the risks remain significant. For example, the Algorithmic Transparency Recording Standard (“ATRS”) is supposed to better ensure the transparency of public sector algorithm-assisted decisions, by recording them in “a complete, open, understandable, easily accessible, and free format”. However, despite being piloted and announced to be compulsory for central Government from February 2024, at the time of writing only 23 completed records have been uploaded.<sup>123</sup>

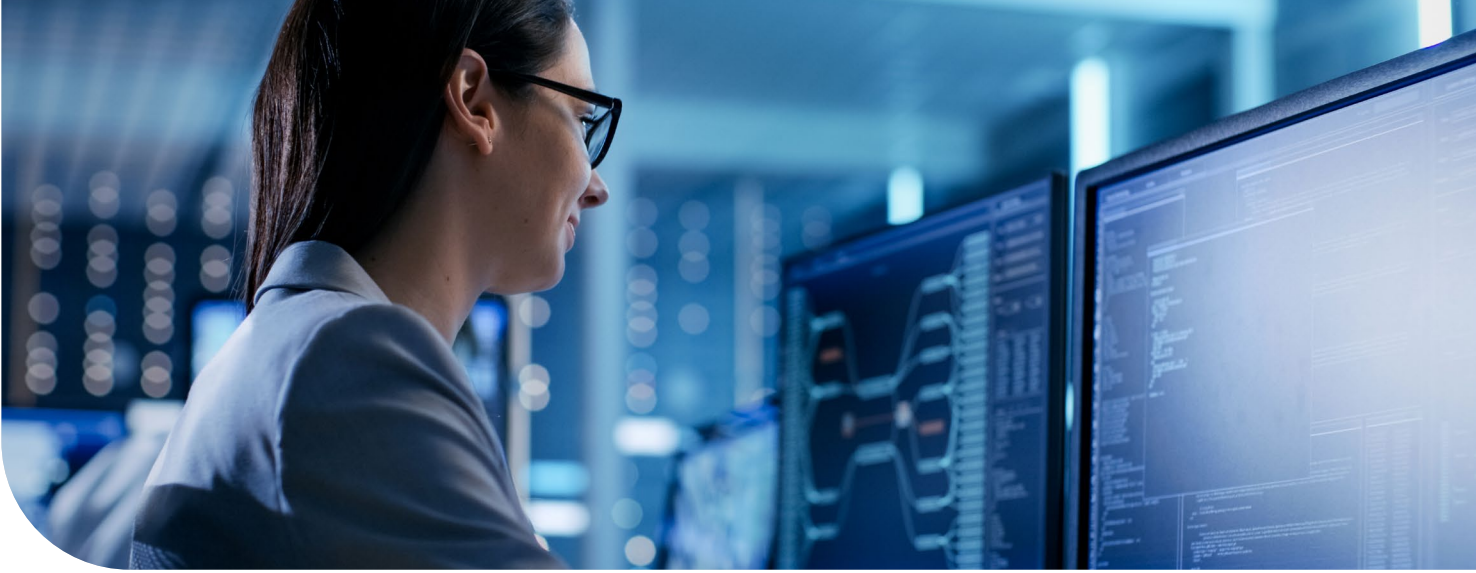


<sup>120</sup> E.g., Karen Waldron et al, ‘From Pilot to Roll Out – Implementing AI Technology’ (Legal Innovators Conference, London, 6 November 2024)—A panellist claimed that Lexis+ did 10 hours of research in 10 minutes; Nick Morgan, ‘How AI is Transforming Due Diligence’ (Legal Innovators Conference, London, 6 November 2024)—Claimed that Xapian AI did seven days of due diligence in seven minutes; Jake Weiner et al, ‘AI in Action: Transforming Legal Practice and Business with Generative AI’ (Legal Innovators Conference, London, 6 November 2024)—A panellist claimed that Harvey AI did a 100 person-days’ worth of document review in three days.

<sup>121</sup> Police and Criminal Evidence Act (PACE) 1984, Code G, para 2.2.

<sup>122</sup> Frank Pasquale and Glyn Cashwell, ‘Prediction, Persuasion, and the Jurisprudence of Behaviourism’ (2018) 68 University of Toronto Law Journal 63.

<sup>123</sup> Department for Science, Innovation and Technology, Cabinet Office, and Central Digital and Data Office, [Algorithmic Transparency Records](#) (2025).



## Lack of accountability for AI decisions

If an individual wishes to challenge the use of AI that has affected them, they must be able to do so practically and effectively. Practical barriers create further risks when AI is involved. For example, when prosecuted in criminal courts by the Post Office, many sub-postmasters were unable to rebut the presumed reliability of computer evidence. Their defence teams could not effectively challenge the computer-generated evidence, resulting in wrongful convictions.<sup>178</sup>

The ability to understand, critically evaluate, and, if necessary, challenge AI outputs also depends on the skills, knowledge, and empowerment of the users – be they decision-makers, professionals, or individuals directly affected by these tools. Even when accessible and effective mechanisms exist, some individuals may lack the skills, knowledge, time, or confidence to challenge the AI tool’s impact on them.<sup>124</sup>

Understanding such complex systems is far from easy. It involves having to understand the purpose of the AI, how it operates, the potential issues which might arise and the application in the specific circumstances. In a fast and emerging field, with increasing leaps in technological capability, these are high bars to meet.

A further risk lies in what is commonly referred to as “automation bias”,<sup>125</sup> the issue of individuals placing too much trust in the outputs of an AI system – and not critically scrutinising its outputs. The impact of this being an increased likelihood of inaccurate or discriminatory outputs going unchecked. This risk intensifies if users lack a clear understanding of what

the AI system is doing, the limits of its capabilities, or the potential issues – such as inaccuracies, biases, or legally impermissible inferences – that may arise. Professional training and guidance can help mitigate automation bias by enabling users to identify errors, verify claims, and confidently reject AI-generated conclusions when appropriate. However, even highly skilled individuals may face environmental pressures – such as high caseloads in lower courts – that encourage over-reliance on AI outputs without adequate scrutiny.

The risks posed are dynamic, in that they differ depending on the context and also the users involved. For instance, a lawyer overseeing the outputs of a legal research tool is far more likely to spot errors and verify information than a layperson untrained in law. This is partly due to the lawyer’s specialised knowledge, but also their access to tools and resources that aid in cross-checking claims. Similarly, users of predictive policing tools should understand that mere AI-driven predictions of criminality cannot lawfully justify the exercise of police powers, which require reasonable, individualised suspicion rather than generalisations or stereotypes. The technical and legal complexity of AI systems demands that users be equipped not only with the right skills, training, and access to verification tools, but also with an understanding of the legal limits of AI outputs. Without such empowerment, individuals are less likely to recognise, question, or challenge problematic uses of AI, further increasing the risk that unfair, unlawful, or discriminatory decisions go unchecked.

<sup>124</sup> These factors have been found to pose significant obstacles to how people self-manage their personal data, Jacob Kröger, Otto Lutz, and Stefan Ullrich, 'The Myth of Individual Control: Mapping the Limitations of Privacy Self-management' (2021) SSRN Working Paper.

<sup>125</sup> For example, see Kate Goddard, Abdul Roudsari and Jeremy C Wyatt, 'Automation bias: a systematic review of frequency, effect mediators, and mitigators' 19(1) J Am Med Inform Assoc 121.



### iii. Systemic risk factors

The impacts of the use of AI within the justice system are not limited to the specific instances of use, but also have the potential for systems wide impacts. Both direct and indirect impacts can flow from the use of AI tools and should be considered when assessing the suitability of a tool in the justice setting. A change in the approach in one area of the system is likely to have a downstream impact on related aspects of the system – understanding these ripple effects is an important part of conducting a full and proper assessment. More broadly there are macro impacts which relate to the overarching functioning of the justice system- examples of this include:

#### Unlawful and arbitrary exercise of powers

AI tools, particularly when adopted without rigorous legal scrutiny or avenues for effective challenge, carry the risk of enabling authorities to exercise their powers in ways that stretch or break the bounds of law. When public officials rely on AI-generated outputs to make decisions without fully understanding or questioning the basis of those outputs, the potential for unlawful or arbitrary action grows. Without accessible mechanisms to contest or review these decisions, individuals may be left with little recourse against AI-driven errors or overreach.

Moreover, the legal frameworks that govern public authority powers often struggle to keep pace with rapid technological advances. AI tools can test the limits of existing law, operating in a grey area where it is unclear whether a particular surveillance method, data analysis technique, or targeted intervention is permissible. For example, the use of facial recognition technology by South Wales Police has raised serious questions about the point at which highly efficient, data-driven law enforcement activities cross over into unjustified intrusions on individuals' rights.<sup>126</sup> As such technologies proliferate, the likelihood of governments or agencies overstepping legal thresholds, or exploiting ambiguities to push their powers further, becomes more pronounced.

#### Stagnating legal development

The law evolves through the scrutiny of evidence, argument, and the testing of principles in new contexts. Courts are not oracles, and case law should not be understood as absolute truth.<sup>127</sup> Judges inevitably make decisions under uncertainty with imperfect information (only as presented by the parties and their legal counsel) and may imperfectly use it.<sup>128</sup> Often, even “reasonable minds may differ on the results of given cases.”<sup>129</sup> Use of AI tools to advise on the law may embed only one interpretation of the law which would otherwise be tested, distinguished or developed. They also risk undermining the individual assessment of each case, leading to generalisations across legal issues.



<sup>127</sup> Holli Sargeant and Måns Magnusson, *Bias in Legal Data for Generative AI* (International Conference on Machine Learning 2024, Vienna, July 2024); Barry Smith, 'Of Truth and Certainty in the Law: Reflections on the Legal Method' (1985) 30 *American Journal of Jurisprudence* 97; Ronald A Heiner, 'Imperfect Decisions and the Law: On the Evolution of Legal Precedent and Rules' (1986) 15 *The Journal of Legal Studies* 227; Jules L Coleman, 'Truth and Objectivity in Law' (1995) 1 *Legal Theory* 33.

<sup>128</sup> Ronald Heiner, *Imperfect Decisions and the Law: On the Evolution of Legal Precedent and Rules* (1986), 233.

<sup>129</sup> Barry Smith, 'Of Truth and Certainty in the Law: Reflections on the Legal Method' (1985) 30(1) *American Journal of Jurisprudence* 97, 119.



## Entrenching discrimination and inequality

At a systemic level, AI tools may deepen existing disparities by reinforcing the social, economic, and structural biases captured in their training data. Over time, this can harden patterns of discrimination into the core of legal processes, making them more difficult to uproot. Instead of enhancing fairness and efficiency, these technologies may solidify the marginalisation of underrepresented groups, perpetuating societal inequalities rather than helping to alleviate them.

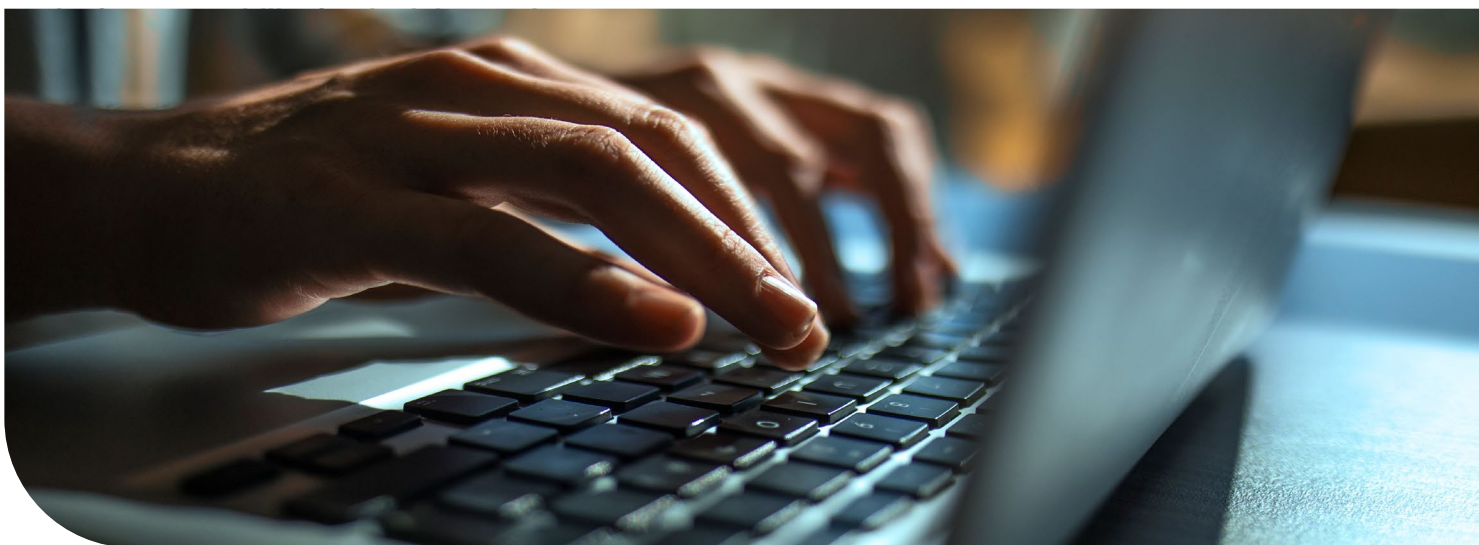
## Unequal access to technology and justice

There is a risk that AI tools which can enhance access to justice will be commercialised and priced in a way that excludes average users.<sup>130</sup> Commercial legal publishers already have been able to establish controlled and monetised access to legal information.<sup>131</sup> This digital exclusion undercuts the principle that justice should be accessible and equitable, depriving less advantaged parties of the analytic precision, strategic insights, and informational advantages AI can provide. Instead of levelling the playing field, the unequal distribution of AI capabilities risks heightening existing imbalances.<sup>132</sup>

AI systems often rely on privately developed technologies, leaving critical aspects of their design and operation beyond the direct purview of public scrutiny. Without adequate contractual or regulatory mechanisms to ensure transparency and responsibility, private vendors and public authorities alike may operate in a space where harm can occur without clear avenues for recourse. For example, such actors may breach human rights but not be considered a public authority under the Human Rights Act 1998 leaving no available recourse. This lack of accountability enables errors and biases to persist, placing fundamental rights at risk without a clearly identifiable entity to hold accountable.

## Eroding public trust and confidence

Finally, the secretive or opaque use of AI can undermine public confidence in the justice system. Particularly, if communities that are already underserved or subject to discriminatory treatment perceive that algorithmic decision-making perpetuates unfair outcomes, trust in the rule of law may erode. As transparency and accountability diminish, the justice system risks being seen as distant, unresponsive, and misaligned with its core mission. Over time, this erosion of public trust weakens social cohesion and undermines the very foundations of a legitimate, democratic society.



<sup>130</sup> Hannes Westermann, 'Using artificial intelligence to increase access to justice' (LLD thesis, University of Montreal 2023); Hannes Westerman and Karim Benyekhlef, 'JusticeBot: A Methodology for Building Augmented Intelligence Tools for Laypeople to Increase Access to Justice' (2023) arXiv:2308.02032.

<sup>131</sup> Natalie Byrom, 'AI risks deepening unequal access to legal information' (2023); Daniel Hoadley et al, 'How Public is Public Law? The Current State of Open Access to Administrative Court Judgments' (2022) 27(2) *Judicial Review* 95.

<sup>132</sup> Natalie Byrom, 'AI risks deepening unequal access to legal information' (2023).

## Case Study: COMPAS

COMPAS, a proprietary recidivism prediction tool used by judges in several US states, exemplifies how opaque, for-profit AI systems can influence critical decisions in the criminal justice system. Developed and sold by a private company, Equivant (formerly Northpointe), COMPAS generates risk scores intended to inform decisions on matters such as sentencing, bail, and parole.<sup>133</sup> Though marketed as a scientifically valid tool capable of assisting judges in evaluating an individual's likelihood of reoffending, independent investigations have cast doubt on its fairness and reliability.

A 2016 review by investigative journalists at ProPublica highlighted troubling flaws in COMPAS's outputs.<sup>134</sup> While the tool claimed to offer objective risk assessments, the analysis found that only around 20% of those predicted to commit violent crimes actually went on to do so. Even more concerning was the discovery that COMPAS was nearly twice as likely to falsely flag Black defendants as future criminals compared to white defendants, reinforcing harmful and long-standing racial disparities in the criminal justice system. These biased predictions risk unjustified harsher treatment for some individuals, undermining the principles of fairness and equality that justice systems are meant to uphold.

Compounding these issues is the lack of transparency: COMPAS's methodology is a trade secret. Judges and defendants are unable to scrutinise how risk factors are selected, weighed, or combined to produce a score. Without knowing which elements drive these assessments, it is difficult—if not impossible—for legal professionals to meaningfully challenge or contextualise the results. The Wisconsin Supreme Court has urged “judicial scepticism” in relation to COMPAS scores, but scepticism alone may not be enough to protect defendants' rights.<sup>135</sup> Judicial independence, impartiality, and competence may be compromised if decision-makers rely on opaque, proprietary systems that cannot be tested or explained, undermining confidence in legal decision-making and the fairness of outcomes.

- **Technical risks:** algorithmic bias and discrimination, uncertain AI methods, lack of transparency of AI methods.
- **Individual risks:** threat to fundamental rights, inability to reverse detriment, lack of transparency of AI decisions.
- **Systemic risks:** Entrenching discrimination and inequality, unlawful and arbitrary exercise of powers, eroding public trust and confidence.

<sup>133</sup> Harvard Law Review, '*State v. Loomis*' (2017) 130 Harv L Rev 1530.

<sup>134</sup> Julia Angwin et al, *Machine Bias* (2016).

<sup>135</sup> *State of Wisconsin v Loomis* 881 N.W.2d 749 (Wis. 2016); K. Freeman, 'Algorithmic Injustice: how the Wisconsin Supreme Court failed to protect due process rights in *State v Loomis*' (2016) 18(5) NC JL & Tech 75.

## Case Study: Predictive Policing

Predictive policing algorithms, such as Geolítica, formerly known as “PredPol,”<sup>136</sup> have been deployed by some police forces to forecast crime trends and direct policing resources accordingly. Drawing on historical data of criminality, these systems claim to enhance efficiency and effectiveness in resource allocation. However, they have been criticised in both the United States,<sup>137</sup> and the United Kingdom,<sup>138</sup> for relying on flawed and biased data that reflects the concentration of police surveillance in racialised communities. As a result, the “predictions” simply mirror existing biases rather than providing a genuinely neutral assessment of risk.

For instance, the use of PredPol in the United States revealed that Black individuals, and those from other minority ethnic groups, were disproportionately targeted. Drug use was found to be evenly spread across the population, yet the algorithm consistently over-targeted communities that were already subject to intensive policing. This not only entrenches discrimination but also creates a feedback loop, wherein historical patterns of biased policing are reinforced and perpetuated over time.

From a legal perspective, such predictive models cannot serve as a lawful basis for exercising police powers: reasonable grounds of suspicion are required, which explicitly cannot be based on “generalisations or stereotypical images that certain groups or categories of people are more likely to be involved in criminal activity”.<sup>139</sup> Without transparency or meaningful opportunities to challenge these algorithmic assessments, individuals face significant barriers to defending themselves against inaccurate or unjustified intrusions on their rights. As a number of stakeholders have observed, there is often scant publicly available information regarding how these risk assessment tools operate and inform decision-making. This lack of transparency can undermine human rights, including the rights to liberty and freedom of movement, particularly if individuals are unaware that they have been subjected to algorithmic evaluation.<sup>140</sup>

- **Technical risks:** Poor or incomplete data, algorithmic bias, uncertain and unpredictable decisions.
- **Individual risks:** threat to fundamental rights, lack of transparency of AI decisions, lack of accountability for AI decisions, inability to reverse detriment.
- **Systemic risks:** Entrenching discrimination and inequality, unlawful and arbitrary exercise of powers, eroding public trust and confidence.

<sup>136</sup> Aaron Sankin and Surya Mattu, *How We Assessed the Accuracy of Predictive Policing Software* (2023).

<sup>137</sup> 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.

<sup>138</sup> Kristian Lum and William Isaac, 'To predict and serve?' (2016) 13(5) *Significance* 14; Miri Zilka, Holli Sargeant and Adrian Weller, 'Transparency, Governance and Regulation of Algorithmic Tools Deployed in the Criminal Justice System: a UK Case Study' 2022 *AAAI/ACM Conference on AI, Ethics, and Society* 880.

<sup>139</sup> Para 2.2B, Code A, the Police and Criminal Evidence Act, 1984, as highlighted in Marion Oswald, 'Give Me a Ping, Vasili. One Ping Only' *Why the Success of Machine Learning Depends on Empowered People*, Crest Security Review (2022).

<sup>140</sup> Australian Human Rights Commission, *Final Report: Human Rights and Technology* (2021).



## Case Study: Facial Recognition Technology

Facial recognition technologies have received significant attention as their use has grown in recent years. Facial recognition technologies are “designed to detect and identify individuals by comparing digital images against a list or ‘database’ of faces. The systems look for “matches” to those in the database. The aim being to assist law enforcement to detect specific people—for instance suspects—quickly, in a way that is not possible by manned patrols especially in large, crowded places.”<sup>141</sup>

Now well-known academic research revealed that facial recognition technology trained largely on white male faces will be more likely to misidentify female, Black and minority ethnic individuals.<sup>142</sup>

As a result, there is a heightened risk of false matches, unjustified police interventions, and discriminatory targeting. In the United Kingdom, the Bridges case—the legal challenge against the use of facial recognition by South Wales Police—highlighted the dangers posed by inadequate safeguards and oversight. Although the Court of Appeal acknowledged the technology’s potential, it found that the police did not have lawful authority to use the tool and had failed to consider or address its potential bias and discriminatory impacts. The Court concluded that too much discretion was left to individual officers about who to place on watchlists and where to deploy the system, creating a risk of arbitrary decision-making and disproportionate interference with privacy rights.<sup>143</sup>

More broadly, concerns also stem from the lack of explicit legal frameworks governing the use of facial recognition technology in many jurisdictions. For example, it has never been explicitly debated in the UK Parliament, raising questions about democratic oversight and the rule of law.<sup>172</sup> EU-level proposals, such as the AI Act, have recognised these issues by restricting the use of live facial recognition and requiring independent oversight, but such measures remain under discussion and have yet to be fully implemented.<sup>144</sup> Without clear statutory guidance, transparent accountability mechanisms, and meaningful opportunities to challenge false matches or biased applications, individuals may find themselves subject to intrusive and unjustified surveillance without recourse.

- **Technical risks:** Poor or incomplete data, algorithmic bias, uncertain and unpredictable decisions, lack of transparency of models.
- **Individual risks:** threat to fundamental rights, lack of transparency of AI decisions, lack of accountability for AI decisions, inability to reverse detriment.
- **Systemic risks:** Entrenching discrimination and inequality, unlawful and arbitrary exercise of powers, eroding public trust and confidence.

<sup>141</sup> Sophia Adams-Bhatti and Holli Sargeant, ‘Algorithms in the Justice System: Current Practices, Legal and Ethical Challenges’ in Matt Hervey and Matthew Lavy (eds), *The Law of Artificial Intelligence* (2nd edn, Sweet & Maxwell 2024).

<sup>142</sup> Joy Buolamwini and Timnit Gebru, ‘Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification’ (2018) 81 Proceedings of Machine Learning Research 77.

<sup>143</sup> *R (Bridges) v CC South Wales* [2019] EWHC 2341; [2020] EWCA Civ 1058.

<sup>144</sup> Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 2024/1689.

## Case Study: Natural Language Processing (NLP)-powered legal research tools

Certain proprietary tools used in the legal industry, such as Lexis+ AI and Westlaw AI-Assisted Research, use natural language processing (NLP) to conduct legal research.<sup>145</sup> These tools can take a variety of inputs, such as user questions and uploaded documents, and output a generated answer. Many, but not all, of these tools use foundation large language models (LLMs) as their base.

These tools have significant potential to advance several goals of the use of AI in the justice system, such as by empowering the competence of lawyers through streamlining the legal research process, and allowing for more efficient delivery of legal services, thereby increasing access to justice.

Current LLMs have a well-documented tendency of "hallucinating," where false or misleading information is produced. There are already documented incidents of false case law being cited in courts across multiple jurisdictions.<sup>146</sup>

As the use of NLP-powered legal research tools gain widespread adoption in the legal sector—with 82% of lawyers in the UK stating that they either use or plan to use generative AI according to a September 2024 LexisNexis survey<sup>147</sup>—it is very credible to predict that there will be large scale reliance by the legal profession on these technology tools. This will undoubtedly bring benefits, as mentioned above, but also increases the risks associated with automation bias and undermine competence and independence within the wider justice system.<sup>148</sup> The increased proliferation of proprietary tools, in particular, can also exacerbate inaccessibility to the law by further entrenching “inequality of arms” between lawyers who can access the tool, and lawyers and litigants-in-person, who cannot.

- **Technical risks:** Hallucination.
- **Individual risks:** Lack of transparency for research, automation bias, individual inequality of arms and/ or digital exclusion.
- **Systemic risks:** Entrenching inequality of arms, eroding public trust and confidence, entrenching discrimination and inequality.

<sup>145</sup> Varun Magesh et al, 'Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools' (2024) arXiv:2405.20362.

<sup>146</sup> Ibid.; E.g., *Parker v Forsyth N.O. and Others* (1585/20) [2023] ZAGPRD 1; *Mavundla v MEC* (7940/2024P) [2025] ZAKZPHC 2.

<sup>147</sup> Dylan Brown, *Future of Law* (2024).

<sup>148</sup> For an expression of similar concerns regarding the use of Google's search engine, see e.g., *S v Aliyu* (A12/2023) [2023] ZAGPJHC 697, para 31.

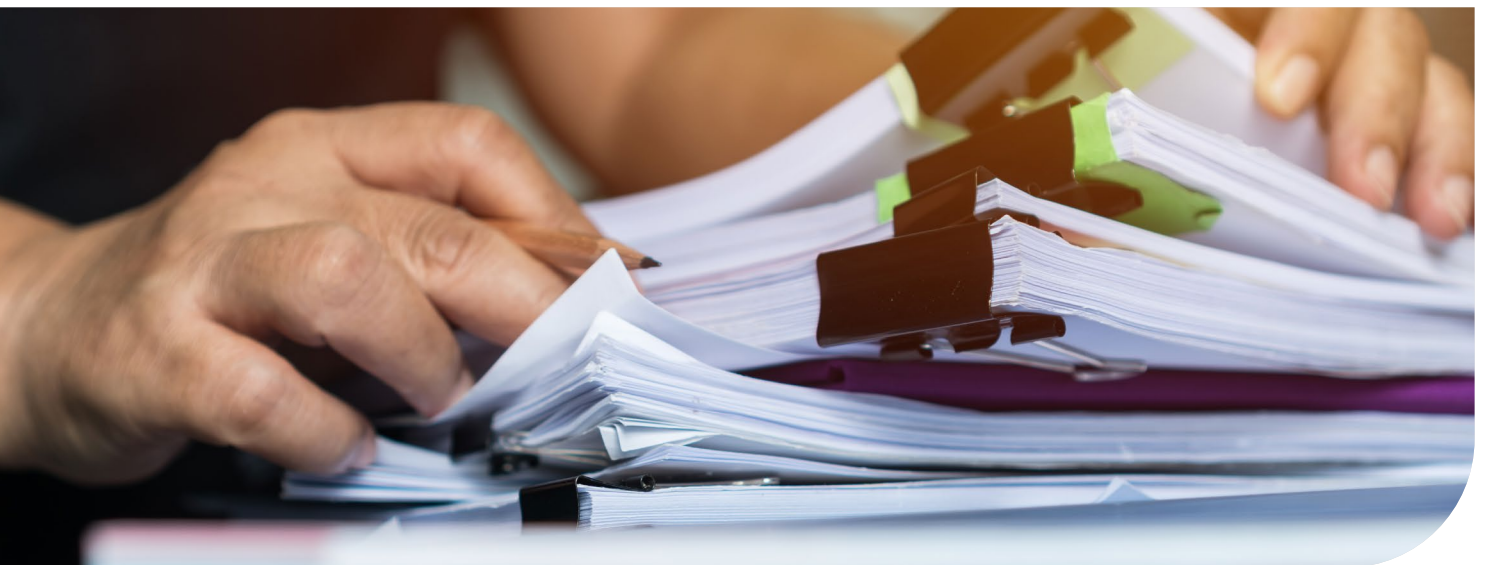
## Case Study: Court transcription software

Courts around the world from Singapore to Lithuania have begun using software that automatically transcribes what is said in court hearings.<sup>149</sup> Many of these tools automatically detect audio input in a courtroom and output transcript texts which can be edited by court staff.

AI-powered transcription tools have the potential to improve access to justice by significantly streamlining a court's transcription process, allowing litigants to receive text versions of court proceedings in a much shorter period of time and greatly reduced cost. It could also improve accountability and scrutiny of the courts and tribunals by providing a comprehensive written record of their activity.

However, there are risks of automation bias, both in relation to poorly operating models and where transcription tools have a high degree of accuracy, leading to the 'rubber stamping' of transcripts with minor but potentially significant errors. Such inaccuracies are not only significant to the specific case but also more widely to the overall competence to the justice system.<sup>150</sup> Additionally, as many of these transcription tools use cloud processing, there is also a technical risk of allowing potentially unaccountable private sub-processors to process extremely sensitive personal data.

- **Technical risks:** Transcription errors, breaches of sensitive or confidential data.
- **Individual risks:** Automation bias.
- **Systemic risks:** Eroding public trust and confidence.



<sup>149</sup> E.g., Singapore Agency for Science, Technology and Research, *Intelligent Court Transcription System* (2022); Agne Limante and Maria Dymitruk, 'AI in Courts in Eastern Europe: Lithuania & Poland' (AI and Technologies in Courts Conference, Zoom, 22 October 2024); Ivana Kunda and Denis Baghrizabehi, 'AI and Courts in Central Europe: Croatia and Slovenia' (AI and Technologies in Courts Conference, Zoom, 22 October 2024); Izuoma Egeruoh-Adindu, 'AI in the Nigerian Court System: Challenges and Prospects' (AI and Technologies in Courts Conference, Zoom, 22 October 2024).

<sup>150</sup> E.g., Ivana Kunda and Denis Baghrizabehi, 'AI and Courts in Central Europe: Croatia and Slovenia' (AI and Technologies in Courts Conference, Zoom, 22 October 2024).

<sup>151</sup> Contend Legal, *Stand up for your legal rights* (2025); DoNotPay, *Your AI Consumer Champion* (2025).



## Case Study: Legal advice chatbots

LLM-powered chatbots such as Contend Legal and DoNotPay take user queries, such as requests for legal advice or requests to draft letters, as input and give generated answers to user queries.<sup>151</sup> These direct-to-consumer chatbots have the potential to significantly improve access to justice by helping large numbers of people at all times of the day and thereby improve fairness in the justice system by empowering people to vindicate their rights.

However, as outlined above, the hallucination risk posed by LLMs could result in the production of poor legal advice or court documents. These risks can be particularly heightened if there are large-scale changes to the law which increases the likelihood of outdated information being produced. Budgetary constraints in the production of these direct-to-consumer chatbots may also mean heavier reliance on sub-processors for cloud hosting and foundation models, heightening technical risks from unaccountable private actors.

- **Technical risks:** Hallucination, breaches of sensitive or confidential data.
- **Individual risks:** Automation bias, undermining of access to justice if individuals are deterred from bringing claims on the basis of erroneous outputs, potential lack of accountability depending on regulation.
- **Systemic risks:** Eroding public trust and confidence, exacerbating inscrutability and unaccountability, exacerbating unlawfulness.

## d. Risk mitigation

The identification and listing out of all the risks above are not an attempt to suggest that none of these can be overcome, and that they should act as a bar to the deployment of AI. Rather it is our view that many can be overcome with careful thought and consideration of the risks and effective mitigation strategies. These mitigations might take longer to develop, they may require higher investment costs at the research and development and deployment stages and honesty and transparency about the limitations of the tool.

For example, some legal technology proprietors have taken different steps to mitigate the risks posed by hallucination of their NLP research tools including:

- **Retrieval-augmented generation (RAG):** Supplementing an LLM's input with data sourced from a pre-existing legal database of human-written case notes, articles, and judgments.<sup>152</sup>
- **Refining decoding methods:** Altering the way in which LLMs generate their output so that its output is less diverse and random, which has been theorised to cause greater rates of hallucination.<sup>153</sup>
- **Restricting generated output:** Ensuring that the end-user does not generate output, but instead links to articles and resources written by other humans (e.g., Lexology's Lexy).<sup>154</sup>

<sup>152</sup> Varun Magesh et al, 'Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools' (2024) arXiv:2405.20362.

<sup>153</sup> Jean Kaddour et al, 'Challenges and Applications of Large Language Models (2023) arXiv:2307.10169, 21-22; 'A Thorough Examination of Decoding Methods in the Era of LLMs' (2024), (2024) Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing 8601, 8605; DeCoRe: Decoding by Contrasting Retrieval Heads to Mitigate Hallucinations (2024) arXiv:2410.18860.

<sup>154</sup> Rico Green, [Lexy is here](#) (2019).

Other mitigations may be relatively straightforward, for example, guidance on appropriate use of AI tools. In the judicial context, the “Artificial Intelligence: Guidance for Judicial Office Holder” sets out key risks and issues associated with the use of AI and provides practical examples for mitigating these.<sup>155</sup>

The development of more robust systems to work alongside and within the existing justice system will in some cases be required to help mitigate the potential harms of AI. In fact, we see some examples of this emerging already. For example the oversight mechanism put in place by West Midlands Police through its Data Ethics Committee provides a useful blue print for how complex considerations can be understood and assessed in the context of front line policing.<sup>156</sup> Or the work of the Online Procedure Rule Committee (OPRC) which is developing rules for online proceedings across the civil, family and tribunal jurisdictions, as well as data and behavioural standards for online dispute resolution.<sup>157</sup>

There is a wide ecosystem of expertise which can be harnessed to try and tackle some of the complex risks we set out, and who can also play an ongoing role in delivering real time independent challenge. Ranging from academic analysis,<sup>158</sup> regulators who can incentivise cultures of best practice, create opportunities to innovate safely, and provide important checks on how standards are maintained, through to public oversight actors – such as Commissioners,

Ombudsmen, Chief Inspectorates who can identify and challenge collective harms – there are many actors who can play a part in mitigating the risks.

This said it is equally true that some of the risks may not be manageable and the impacts too important to ignore, and in those cases, we argue that the imperative should be to protect the rule of law and human rights, and to draw a red line until the issues and concerns can be resolved. As things currently stand, as identified throughout this report, there are many factors which make managing the risks we have identified above incredibly hard in some cases. The justice system has more data gaps than any other public service,<sup>159</sup> limiting the extent to which AI used in the justice system can be monitored for its impact. There has been a decline in court reporting,<sup>160</sup> “discord and dysfunction” in the regulation of legal services,<sup>161</sup> an unclear, overlapping, and uncoordinated regulatory landscape in criminal justice,<sup>162</sup> and failing regulation of equality and human rights in need of refocus.<sup>163</sup> Contract management when working with private companies in the justice system has been subject to concerns as we have previously reported,<sup>164</sup> and the distinction between “legal advice” and “legal information” is becoming increasingly blurred with technological advancements,<sup>165</sup> giving rise to potential gaps in consumer protection.<sup>166</sup> These factors are mere illustrations of the real-world context in which the complexities of managing the risks of AI development and deployment in the justice system is incredibly hard.



<sup>155</sup> Courts and Tribunals Judiciary, [Artificial Intelligence \(AI\) Guidance for Judicial Office Holders](#) (2023)

<sup>156</sup> Marion Oswald et al, [A preliminary study of West Midlands Police's specialist data ethics review committee](#) (2024), 106.

<sup>157</sup> Ministry of Justice and Christopher Ballamy, [New Online Procedure Rule Committee launched](#) (2023).

<sup>158</sup> E.g., Varun Magesh et al, ['Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools'](#) (2024) arXiv:2405.20362.

<sup>159</sup> Thomas Pope, Gavin Freeguard, and Sophie Metcalfe, [Doing Data justice: Improving how data is collected, managed and used in the justice system](#) (2023).

<sup>160</sup> House of Commons Justice Committee, [Open justice: court reporting in the digital age](#) (2022).

<sup>161</sup> House of Commons Justice Committee, [Justice Committee sets out recommendations on the regulation of the legal professions to the Lord Chancellor](#) (2024).

<sup>162</sup> Ibid.

<sup>163</sup> House of Commons Women and Equalities Committee, [Enforcing the Equality Act: the law and the role of the Equality and Human Rights Commission](#) (2019).

<sup>164</sup> JUSTICE, [Beyond the Blame Game](#) (2024).

<sup>165</sup> Judith Bennett et al, [Current State of Automated Legal Advice Tools](#) (2018) RALAT Working Paper No 1; Keith Porcaro, Keith Porcaro, ['Gray Advice'](#) (2023) 25 Duke Law & Technology Review 48.

<sup>166</sup> A notable case in the United States highlighted this distinction, where it was found that a lawyer conducting document review was not engaged in the “practice of law,” and tasks performed entirely by machines cannot, by definition, involve legal judgment. *Lola v Skadden, Arps, Slate, Meagher & Flom LLP* (2015) 620 Fed Appx 37 (2nd Cir).

# 5. Conclusions





AI has the potential to be a game changer in many areas of society. The justice system in England and Wales, (and in many other jurisdictions) is struggling to keep pace with what is needed of it. On the face of it we have a perfect match – a system in need of dire help without placing greater pressure on the public purse and an incredibly useful general-purpose technology which can be deployed from front line policing, through to prioritising court time, or assessing recidivism risks. However, the justice system is not comparable to other aspects of everyday life – buying groceries, choosing clothes, or podcast preferences – it plays a critical and keystone role in the functioning of democratic societies.

For this reason, the deployment of AI in the justice sector should be the result of careful consideration, and one guided by a clear purpose and responsibility. At the heart of this purpose is the embedding of the rule of law and the human rights framework as the guiding principles. Our framework, which seeks to embed the rule of law and human rights, sets out a clear pathway to assessing the suitability of any given AI system for use in the justice sector.

Firstly, does the proposed system or tool advance any of the goals of justice system? In other words, does the AI help improve the effectiveness of access to justice? Does it help to deliver lawful and fair decisions? Does it help to bolster trust and confidence through openness? If the answers to all of these questions is no, it begs the question what is the value of the AI to the justice sector?

Secondly, even with a clear purpose which supports the advancement of the justice system, delivering on that purpose is a different challenge altogether. There are technical challenges to overcome – such as poor data – and wider complex policy considerations. At the core of managing these issues is the responsibility – at every stage from design through to deployment – to be fully aware of the implications, impacts and risks and to mitigate those risks. If such risks cannot be mitigated sufficiently, those involved in the proposed deployment must be willing to go back to the 'drawing board'.

This framework is applicable to any area of the justice system – corporate settings, civil and family courts, criminal through to administrative back-office functions. We urge policy makers, developers, purchasers and users to adopt this framework to inform decision-making and improve current practice. Additionally, in a sector where investment in AI development is increasing, having a unified high level policy framework which spans all areas of the justice system has the benefit of making sense out of a very 'noisy' arena with competing and somewhat confusing approaches. This lack of coherence serves both the justice system and the innovation sector poorly.

It would be naive to suggest that this one policy framework is sufficient on its own. It is for this reason that we propose to develop, in partnership with the sector, deep dive analyses of each practice area, exploring the application of this framework in each setting and draw out specific issues and potential solutions. We welcome inputs to this work from actors across the justice system with the desire to positively find ways to deliver a justice system which meets the needs of the public.



# Annex 1: Illustrative Framework Assessment Tool

Understanding the purpose of the AI tool, the implications for the rule of law and human rights, considering properly the impacts – both positive and negative should be part of the norm for anyone designing, developing, or deploying AI in the justice sector. From data scientists to policy makers in Whitehall there is a duty to act responsibly.

To help guide initial analysis we have set out a number of prompt questions which can be used to initiate relevant discussions.

Each one of these questions should be regarded as a prompt to better understand the AI being developed or used. These questions are not proposed to be used as a binary scoring matrix.

1. What problem is this tool seeking to deal with?
2. With the justice-goals in mind, which of these 3 goals does this tool seek to assist with?
3. What are the proposed benefits?
4. What are the risks – technical, individual and systemic?
5. How do these benefits weigh against the risks?
6. Can these risks be mitigated? If so, how?
7. Are the residual risks acceptable?
8. What assumptions is the risk model based on? How secure are these assumptions?
9. What are the impacts of the residual risks at both an individual level and systems wide?
10. What are the limitations of the tools? And in what circumstances is the tool not suitable?
11. How will risks, mitigations, benefits be monitored and evaluated?
12. Are other strategies (AI or otherwise), including doing nothing, better?

# Acknowledgements

I would like to thank JUSTICE's Senior Lawyer **Ellen Lefley** who led the work on this report for JUSTICE and **Dr Holli Sargeant** for her invaluable research and drafting support. We are also very grateful to JUSTICE's Linklaters fellows, **Daniel Leung** and **Nandini Mitra**, and to **Chris Martin** and his team at **Pinsent Masons LLP - Concetta Dalziel** and **Charlotte Bowen** – for their research assistance.

We are incredibly grateful to all those who spoke with us to share their expertise. Particular thanks to our group of consultees who provided guidance and feedback on the report: **Orlando Conetta** (Pinsent Masons), **Professor Marion Oswald** (Northumbria University and The Alan Turing Institute), **Louise Hooper** (Garden Court Chambers), **Professor Felix Steffek** (University of Cambridge) and **Matthew Lavy KC** (4 Pump Court). The final report does not reflect these consultees' individual positions nor those of their respective institutions and affiliations.

Thanks to **Callum Brown** (Simmons Wavelength) and **Nerys James, David Pegg** and **Robin Hill** (Oliver) for their design assistance in producing the report.

And finally, thanks to **Hayley Sayers, Meghan Higgins** and **Alice Perry** at **Pinsent Masons LLP** for supporting the production of this report from its conception to its launch.



**Sophia Adams Bhatti**





