

# Chapter 14

## Automated Justice: Issues, Benefits and Risks in the Use of Artificial Intelligence and Its Algorithms in Access to Justice and Law Enforcement



**Caroline Gans-Combe**

**Abstract** The use of artificial intelligence (AI) in the field of law has generated many hopes. Some have seen it as a way of relieving courts' congestion, facilitating investigations, and making sentences for certain offences more consistent—and therefore fairer. But while it is true that the work of investigators and judges can be facilitated by these tools, particularly in terms of finding evidence during the investigative process, or preparing legal summaries, the panorama of current uses is far from rosy, as it often clashes with the reality of field usage and raises serious questions regarding human rights. This chapter will use the Robodebt Case to explore some of the problems with introducing automation into legal systems with little human oversight. AI—especially if it is poorly designed—has biases in its data and learning pathways which need to be corrected. The infrastructures that carry these tools may fail, introducing novel bias. All these elements are poorly understood by the legal world and can lead to misuse. In this context, there is a need to identify both the users of AI in the area of law and the uses made of it, as well as a need for transparency, the rules and contours of which have yet to be established.

**Keywords** Artificial intelligence · Law enforcement · Judicial system · Evidence gathering · Sentencing · Adversarial right · Automation · Enquiries · Overcollection

When Deanna Amato had her tax deductions withheld by Centrelink in 2019—a service of the Australian government's Income Compliance Program, responsible at the time for, among other things, the recovery of social security overpayments—she had no idea that this was one of the many erroneous outputs of an automated legal decision-making software. The so-called Robodebt case involved hundreds of thousands of Australians being issued automated reports that they owed money to the Australian welfare system. It was later acknowledged that the use of inadequate

---

C. Gans-Combe (✉)

Insec Business School, Head of Structured Research, Omnes Education Research Center, Paris, France

e-mail: [cganscombe@omneseducation.com](mailto:cganscombe@omneseducation.com)

© The Author(s) 2022

D. O'Mathúna and R. Iphofen (eds.), *Ethics, Integrity and Policymaking*, Research Ethics Forum 9, [https://doi.org/10.1007/978-3-031-15746-2\\_14](https://doi.org/10.1007/978-3-031-15746-2_14)

175

data associated with a rogue algorithmic practice resulted in the Australian Government sending out more than 400,000 requests for refunds of non-existent social overpayments (Bennett 2019).

This so-called “Robodebt” case demonstrated that the uncontrolled use of artificial intelligence and machine learning tools for decision-making purposes with legal impacts is far from mature, notwithstanding the billions of dollars in revenue already generated annually by these types of tools (Alston 2019).

***The Robodebt case:*** The Robodebt scheme, also known as the Centrelink debt recovery program, ran between 2015 and November 2019. Using comparison algorithms, Centrelink, an Australian social security state agency, matched the average annual tax rates and tax payments of tax deduction recipients with their actual fortnightly tax returns. If a difference in amount was found in favour of the social security administration, the algorithm systematically concluded that there was an error in the declaration and automatically launched a recovery procedure. In effect, welfare recipients were assumed to owe money until they could prove otherwise. Even worse, Centrelink issued a debt notice based on averaged Australian Taxation Office (ATO) information alone and applied a 10% penalty. Centrelink did not use its information-gathering powers to contact the concerned taxpayer’s employers or banks to provide details of actual earnings for the relevant fortnights to determine whether the taxpayer actually owed the automated debt calculation. When exposed through the Amato case, the proceeding resulted in a large class-action suit, exposing a massive failure of public administration as the government unlawfully raised \$1.76 billion in debts against 443,000 people. During that time, the Australian government had also pursued about 381,000 people, unlawfully recovering \$751 million, including through private debt collectors. In court, the Australian Government conceded that the averaging process using ATO income data to calculate all Robodebt was unlawful.

Thus, the illusion that administrative or legal procedures can be facilitated or accelerated clashes with the reality of field usage, and raises serious questions regarding human rights (Langford 2020), notwithstanding the strong impacts on involved or concerned parties and an even more significant effect on vulnerable populations (Carney 2020).

In this context, this chapter has three objectives:

- To identify the types of cases concerned by these situations to propose perimeters of vigilance as well as to identify the critical paths leading to potential errors.
- To identify (through the different cases raised in existing publications) the elements of technological design that can lead to the scandals and biases mentioned above: the “glitches” in the sense of Meunier et al. (2019).

- To suggest some good practices to help the supporters of this type of solution to prosper without damaging public liberties, the rights of the citizens or the possibilities of unbiased access to law and justice.

## 14.1 Where and How the Automation of Legal Acts Takes Place

The example of “Robodebt” has made it possible to highlight that the use of algorithms in processes with legal impacts is not limited to the courts of justice. Artificial intelligence (AI) is the set of theories and techniques implemented to create machines capable of simulating human intelligence in its neural architecture for various uses and in various fields. It combines a set of technological tools (deep learning, structured or unstructured predictive analysis, data) at the origin of methods for solving problems with high logical and/or algorithmic complexity. There are therefore many uses open to AI (support, data collection, decision making, learning, prediction), which can be found in what is known today as LegalTech, i.e. the involvement of technology in developing, proposing or providing products or services related to law and justice. The LegalTech market will be worth more than US\$25 billion by 2025, with US\$6.37 billion in the European market alone (Alsop et al. 2020).

It, therefore, seems useful, as a preliminary step, to propose a mapping of the concerned domains. The mapping proposed in this chapter is matrix-based. First, it is divided into two branches, the branch where the action of algorithms replaces the judge and the branch where the action of algorithms supports the judge, each of them having its own justifications and operationality, and raising questions about legality, the legal system, public liberties and even efficiency. Finally, it is grouped in a transverse questioning that covers what could be called predictive or prophetic law: the data coming from both the results of the decision support and the positive actions of substitutions to the judge (for example, which automated decisions are more likely to be appealed) being themselves concatenated to permanently reformulate the perimeter and the distribution of the elements of the matrix (Fig. 14.1: The distinctive branches of AI use in the field of law and justice).

Today the substitution—at least the one perceived by the simple citizen—often covers only subjects of first instance law (parking fines, speeding tickets, tax or social deductions), but tomorrow we can imagine a tightening of practices limiting the appeal procedures or making them as complex as possible to avoid the occurrence of such disputes. The practical result would be the use of algorithms to distance the citizen from the law or from the exercise of his rights. The appeal of an automated decision is indeed quite complex, especially to access a human explaining clearly the procedures to follow (and even when he is trained to do so). In the Robodebt case, the Centrelink service at the origin of the scandal had such a long waiting time (more than 15 min online) to access a call center of untrained agents, that the only recourse for the citizens was a class action suit which still has a devastating impact both financially and in terms of image for the company, notwithstanding the intervention of many



**Fig. 14.1** The distinctive branches of "AI uses" in the field of law and justice

spin doctors to try to redress the situation (Towell 2017). Indeed, the proceedings, some of which are still ongoing, have so far cost the Australian state and the social security administration a total of A\$1.8 billion.

The substitution of judges by other legal operators, such as arbitrators or mediators, called alternative disputes resolutions (ADR), is becoming increasingly common, in business or family law for example. There is thus a habit of differentiated practices that is taking root in society (Ferrand 2015).

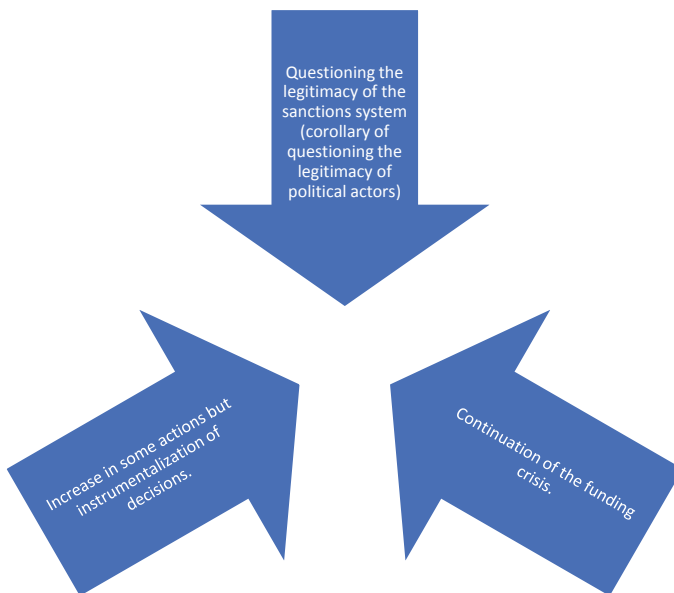
At the same time, the interference of digital technologies in the legal process is not limited to the use of artificial intelligence (Ulenaers 2020). It is an older and broader issue (Stranieri et al. 1999) than the question underlying the intrusion of this new approach into the legal world or in the service of the legal ecosystem. Therefore, the issues raised are both non-specific and specific to AI. All the subjects to be analyzed below related to access to law, and to the respect of the adversarial process, are non-specific inasmuch as they relate to automation as to AI, whereas AI raises specific

issues in the approaches related to network and data architectures, to the quality of the data and especially its neutrality. These new uses (deep learning, etc.) increase the paradox between the expectation of neutrality with regard to the judge's decision and the respect for public liberties.

### 14.1.1 Why Substitute Judges with Algorithmic/Automated Operations?

The substituting of judges with algorithms has been justified by the need to relieve overloaded courts of justice. Political actors on all sides have been trying to do this since access to the law was considered a fundamental right during the Enlightenment (Gray 2008).

Indeed, access to the judge to settle disputes with the constituted bodies (police, state, etc.) grew very quickly as people developed an expectation for a justice system that was autonomous from political actors (Bell et al. 2004), and that could be considered neutral. However, since the beginning of this century, justice, like the political-administrative ecosystem that is part of the rule of law, is in crisis (Stoett 2019). It suffers from a triple conflict between expectations, means and legitimacy (Fig. 14.2: The three tensions facing the judiciary).



**Fig. 14.2** The three tensions facing the judiciary

*Expectations*: if recourse to depenalization—i.e. the exclusion of certain disputes from the scope of the first instance judge—has had a definite effect on the volume of cases to be tried (we will come back to this in detail below), the contestation of the administration’s actions and decisions in courts is at best stable, and at worst rising sharply, including in member countries where the rule of law is undermined.

*Means*: if the financial means are globally increasing, they remain quite scarce compared to the populations concerned: in this respect, Switzerland (a non-EU member state) is by far the European country that invests the most money per inhabitant in its judicial system. With an expenditure of €220.6 per inhabitant in 2018, well ahead of Luxembourg (€163.5) and Germany (€131.2). The three least well-endowed states within the European Union are Lithuania (€41.5 per inhabitant), Bulgaria (€42.3) and Romania (€42.6). All other EU members fall in between (European Commission 2021).

*Legitimacy*: the imperative of judicial neutrality (Ellis 2018) has been undermined on the one hand by the instrumentalization of the processes of appointing magistrates by politics, and on the other hand by the politically marked decisions that have quite naturally followed (Delsol 2021).

The impact of this probing of the legitimacy of decisions taken by judges has been reinforced by the relatively continuous questioning by whole segments of the European population—since the beginning of the COVID-19 crisis—of the legitimacy of elected decision-makers. When these actors have not demonstrated what Parviainen et al. (2021) have called “Epistemic Humility”, problems arise. Instead, they present to the public their sometimes biased certainties. “When the law obliterates itself at the top, everything becomes permitted,” Jean Pierre Delsol (2021) rightly tells us.

To *a request for neutrality*, the public actors answered by the automation of certain decision-making. This was based on the following presupposition: A decision made by an automated device would not be influenced by humans and would be in compliance with the law. This is true, but a decision that is legal at a given time does not imply that it is right or fair (Barocas and Hardt 2017), nor that this legality will remain over time. History is replete with the instrumentalization of the rule of law for political purposes—is it worth recalling the sinister legalism of the Nazis who made the Final Solution legal under the laws of the time and the country, of which Hanna Arendt has painted a vivid picture (Poizat 2017)? Robodebt, or the fines imposed on drivers who respect the law, in this case, the speed limits—due to a change in legislation not carried over in time in the machines—underlines the fragility of the assumption that the machine can’t go wrong. Yet that assumption is still supported by many public actors in Europe (Pasquale 2019).

These are the first limits to legal automation: the hiatus between the legality of the sanction and the reality of the infringement. Just because an offence has been found by a machine does not mean that it is genuine or meaningful from a legal point of view. The context is just as important as the offence itself.

The appetite for neutrality appears therefore as one of the elements of the observed tendency to automate judicial proceedings. Other more fundamental intentions are

added to this. It will be recalled that the European Union measures the legal performance of each member country according to three criteria (European Commission 2021):

- **Efficiency**, which focuses on indicators such as the length of proceedings, the rate of variation of the stock of pending cases and the number of pending cases.
- **Quality**, which highlights indicators related to accessibility, such as legal aid and court fees, training, budget, human resources, and digitization.
- **Independence**, which focuses on indicators relating to the perception of the independence of justice by the general public and businesses, guarantees concerning judges and guarantees concerning the functioning of national prosecution offices.

## 14.2 Problems Raised by the Substitution of Automation for the Judge

Admittedly, as indicated in the latest iteration of the Council of Europe’s European Commission for the Efficiency of Justice (CEPEJ 2020a), automation makes it possible to reduce the response time to a legal problem, and thus the congestion of the courts, especially when they deal with only relatively low-value disputes. But this practice poses serious problems as regards the rule of law because:

- a. this is problematic in terms of respecting the principle of adversarial proceedings, which is a positive right recognized both by Article 47 of the Charter of Fundamental Rights of the European Union and by Article 6 of the European Convention on Human Rights.
- b. it requires the removal from the control of the judge (for the benefit of administrations often under the tutelage of the Executive) a whole part of the societal activity by decriminalizing certain facts (Croze 2017).
- c. it condemns more than it exonerates. The automation is systematically incriminating because it is only interested in the potential offender, but it can be wrong (we have seen this with the problems of speeding or parking tickets as well as Robodebt). The burden of proof is then on the potential offender, who must prove that he was not at fault, which is complex, since he may not have legal training (Maurel 2019). The wrongdoing is then converted into a sanction without measure or understanding of the ecosystem. As such, the process breaks the social dimension of justice by reducing it to a mere decision. Justice is fundamentally a process, and trials, as they unfold, contribute to the acceptance of sentences.
- d. it raises questions about access to the law and corrective procedures in the event of a challenge to the ultimate decision.

It should be recalled that the European Commission itself, in order to deny access to documents and elements of internal procedures, hides behind the notion of “space to think”. The legal contours here are vague, even though the principle is expressly condemned by the Court of Justice of the European Communities (Judgement of the General Court (Seventh Chamber, Extended Composition) of 22 March 2018 in case

T-540/15 *De Capitani / Parliament*) on the grounds that an EU decision cannot be opposed to a citizen on the basis of elements to which he or she has not had access (Hillebrandt and Novak 2016). “It would violate an elementary principle of law to base a judicial decision on facts and documents of which the parties or any of them have not been able to take notice and on which they have therefore not been able to adopt a position” (Judgment of the Court of 22 March 1961.—*Société nouvelle des usines de Pontlieue—Aciéries du Temple (S.N.U.P.A.T.) v High Authority of the European Coal and Steel Community*.—Joined cases 42 and 49/59).

The principle of adversarial proceedings has been consecrated as a general principle of law by most of the supreme national jurisdictions in European countries (Ferrand 2000) and is included in Article 6 §1 of the European Convention for the Protection of Human Rights (CESDH). Even if this principle seems to apply to only the judge according to the drafting of the aforementioned Article 16, such is not the case since all the legal actors are also debtors of the respect of the contradictory. The European Court of Human Rights (ECHR) stated this in its *Mantovanelli v. France* judgment (ECHR, March 18, 1997, no. 21497/93). In this case, the applicants challenged the enforceability of the expert report on the grounds that the conclusions of the legal expert were based on elements that had been communicated to them without respecting the adversarial process.

The adversarial principle guarantees that the parties will not be judged without having been heard, or at least called. A person who has not been informed of the proceedings against him or her has certain guarantees, both in terms of the remedies available and the enforcement of the decision. The principle of adversarial proceedings guarantees each party the right to be informed of the arguments of fact, law and evidence upon which it will be judged. The various participants in the trial must therefore be loyal and diligent in communicating their documents and conclusions. Any element produced in court must be able to be debated and must therefore be communicated to the opponent.

The judge himself is required to respect the adversarial principle. For example, when he considers raising a legal argument of his own, he must give the parties the opportunity to explain this point, otherwise, he will not be able to use it in his decision. In the *Mantovanelli v. France* judgment (ECHR, March 18, 1997, no. 21497/93), the ECHR recalled that with the principle of adversarial proceedings, “the essential thing is that the parties are able to participate adequately in the proceedings before the ‘court’.” “None of this is possible in the case of automated decisions because the latter precede the debate, not follow it.

There is thus potentially a pronouncement of the sanction “in absentia”, based on exogenous mechanical elements (the speed is  $x$  and such and such a driver has exceeded it by  $y$ ) without the litigant having any means of ex-ante contradiction. The context of non-compliance with a law is not addressed: should a driver be convicted of speeding when he could potentially be transporting a dying COVID-19 patient to the hospital? A judge would admit the (so-called) mitigating circumstance and probably limit the impact of a conviction to the pedagogical understanding that not respecting road safety constitutes a danger. The machine will automatically condemn, and make the person pay a fine before allowing the launching of a possibly



contradictory procedure. A deposit is often compulsory to contest the case, which restricts access to the justice system for the poorest people who do not have the means to block the funds or to mobilize the means to take legal action.

An automation process, or an AI that would respect this principle, has yet to be invented. Therefore, the solution currently found is the depenalization of certain situations. “Legal depenalization” covers two hypotheses. Firstly, it can consist of a downgrading of a given behaviour within the penal system: a crime becomes a misdemeanour; a misdemeanour becomes a contravention. The behaviour remains prohibited but punished less severely. Secondly, depenalization can also consist in taking a behaviour out of criminal law and into another legal sphere: civil law, administrative law, etc. (Jaafari 2016). The said conduct is still prohibited, but it will be sanctioned by another authority such as an alternate court circuit, or by an independent administrative authority such as the Competition Authority in France (financial sanction: Coulon 2008). Depenalization is frequently used to transform road traffic offences into simple contraventions (Fallery 2019).

Since the object is no longer the domain of the judge but of an administrative authority, not only does automation become possible but the nature of recourse to the judge is displaced: it is no longer a question of validating the existence or not of a fault but of arbitrating between two interpretations of a fact by two actors who are opposed to each other and whose relations are necessarily unbalanced (Ellis 2018). It is no longer the triptych ‘law—judge—citizen’ that meets but the triptych ‘citizen—administration—law’ arbitrated by the judge. The law is set by the administrative authority and the judge must only validate the adequate interpretation. The question put to the court is no longer, ‘is there fault therefore sanction?’, but, ‘is the sanction valid with regard to the offence?’.

In this sense, automation transforms profoundly and durably the balance of the social interactions between litigants and justice, at the risk perhaps of upsetting such interchanges. Indeed, in this automatic response, there is a break between cause and effect, and the person subject to trial who has not been confronted with justice will probably find it difficult to understand what he or she is accused of, and therefore to accept the sentence (Bathae 2017). This is another element feeding the crisis of confidence that citizens have towards their institutions (Guggeis 2020).

In this context, artificial intelligence—and deep learning—is one more tool at the service of an already advanced process. It is not so much the process of automation that is being questioned, but how the architecture of AI and associated technologies impact existing automation tools by modifying them, even though the legal questions they raise have hardly been resolved.

This inquiry makes sense when we look at the second branch of our mapping of automation in the field of justice (Fig. 14.1): the support provided to the judge and to the judicial process.

### 14.3 Automation as a Support to the judge's Decision or as a Procedural Assistant

Another area where automation is nowadays omnipresent is judicial decision support. The issue here is far more critical, as it is not limited to justice alone, but also to investigative procedures; in other words, to law enforcement.

The Council of Europe has since 2018 deployed an ethical charter of good practice in the use of AI in legal proceedings and/or as an investigative tool. This doctrinal approach rather than positive law is to be regretted, as it singularly weakens the scope of this publication, which therefore has no binding value (Hyde 2019).

Legal proceedings, especially in criminal matters, call for much information from various sources and of various kinds. The role of AI in legal matters (Prakken and Sartor 2002) is to classify, link and integrate them without omitting any elements. With this support, the investigator and/or the magistrate can view, analyze and compare more documents in an increasingly voluminous and complex file.

AI is a complex set of means likely to endow computer systems with cognitive capacities in order to analyze data in large quantities, with statistical and probabilistic tools and with algorithms that classify, value and confront these sets.

Obviously, there are access limitations related to confidentiality, security or anonymization. Nevertheless, these data can naturally be coupled with other data, which may or may not be freely accessible, in order to constitute resource bases for end-users capable of linking, classifying, contextualizing and cross-referencing them. This allows finer information, with a better granularity, to be calculated or extrapolated than exists in their sources.

In this complex context, two “applications typologies” exist: expert systems and machine learning. In expert systems, the tool is taught by providing it with examples; in machine learning, the tool is asked to find, among the data it has access to, what is likely to be of interest. Expert systems are built on explanatory models to which parameters must be associated to establish one or several results. Machine learning algorithms are used in situations where a mechanistic approach not enhanced by the performance of algorithms is to be excluded.

Hence, machine learning will aim to develop a realistic correlation model between predictive variables (input data) and target variables (results). To do this, machine learning relies on a set of statistical tools and computer algorithms automating the construction of a prediction function (Singh et al. 2007).

An expert system is a tool capable of replicating the cognitive mechanisms of a human expert. For example, it could be software capable of answering questions, by reasoning from known facts and rules. It has 3 parts: a fact base, a rule base and an inference engine: a “thinking machine” (Ross 1933). *The inference engine can use facts and rules to induce further facts until it finds the answer to the given “expert query”.*

Machine learning uses an inductive approach to build a mathematical model from data, including many variables that are not known in advance. The parameters are configured as they are learned in a training phase, which uses training data sets to find

and classify links. Designers chose different machine learning methods depending on the nature of the task at hand.

Expert systems rely on mathematically constructed logic and use deductive reasoning. As such, they are considered the first level of what constitutes Artificial Intelligence. Despite their performance in many fields, the choices deduced by the algorithms are difficult to explain in simple terms. This difficulty is increased by the paradigm shift introduced by deep learning. Thus, the choice of the learning technique used, the ethics of design, the absence of bias and the absolute necessity of being able to explain the arguments that allowed the tool to make a particular decision are all prohibitive elements in the initial construction of these support tools. This is where the problems arise (Whittlestone et al. 2019).

Several emblematic cases can be mentioned:

### ***14.3.1 Easing Access to Law Enforcement and Filing***

The deployment of automation and AI speeds up how files are processed, sorted, and considered, contributing to access to the law. Professionals see their workload reduced by tools like “telerecours” (facilitation of procedures reserved for legal professionals), while other instruments bring citizens closer to law enforcement authorities, notably through automated complaint tools (for example, with cyber-crimes: <https://www.europol.europa.eu/report-a-crime/report-cybercrime-online>).

### ***14.3.2 Solving Criminal Cases: Artificial Intelligence and Judicial Investigation***

During an investigation, data are linked together with observations of a crime scene or misdemeanour. In a few years, this has gone from sketches to 3D extrapolations which include the statements of traces and various objects whose utility for a file is not necessarily established at the very beginning of the enquiry. We are in the middle of over-collection of declarations, testimonies, and hearings which often constitute the first step in the search for contradictions or verifications of facts. However, this collection is a prerequisite for the establishment of what is known as the index value, i.e. the interest and justification of the samples. If AI arrives to accomplish classification (and thus of operational enhancement through the construction of logical links), cross-checking and connecting facts and clues in the judicial files, the expectation is that the truth will be exposed in the trial. For this, various methodologies are used ranging from systematic automation (such as for recognition of people or objects), the establishment of models of existing relationships between information (so-called relational links), the detection of inconsistencies, for example between testimonies using semantic tools (Keyvanpour et al. 2011). All this information is made available

to the investigator or the judge without any filter and often without him understanding its full power and scope. Of course, the demonstration of the truth implies authorizing a certain form of exploitation of mass data but ignores ethical questions about over-collection, post-processing, extensive cross-referencing, etc. If the matching work is well understood and structured, the question of false positives remains open. The balance between revealing the truth and considerable means of investigation must be raised. For if these practices are undeniably useful, they require an objective and strong framework. Otherwise, they risk becoming social scoring or the involvement of innocent third parties whose only fault will have been to be virtually identified as having a link with a person of interest.

In this sense, the data collected is central, but also, the data architecture, particularly about the links established. It is certainly necessary to control the algorithms, their level of transparency, but probably also the objectives given to them, because if not done, the power of AI could be misused. For example, tools could anticipate jury decisions by analyzing their profiles (especially in the United States, where contradictory opinions are expressed, including during jury selection). This differs from France, where juries are drawn by lot from citizens, and are used less and less in favour of professional judges. Another example could be the anticipation of judges' decisions by analyzing their pronouncements. Depending on the analytical results, legal strategies may evolve where some litigants may request a change of jury or of location, while others will not have this possibility (Thagard 2004). Similarly, the use of AI attacks the integrity of the evidence by restricting the judge's choice over the validity of the expert opinions provided (Katz 2014). Such examples call for the greatest caution in weighing the contributions of this type of tool to the judicial world.

### ***14.3.3 Predictive or Prophetic Justice***

We must also mention areas that carry the most fear and fantasy about automation: predictive or prophetic justice (Queudot and Meurs 2018). This involves predicting the future of a legal action in order to anticipate it (which is sometimes presented as risk management) or to harmonise the scope of sanctions according to given offences/crimes. In addition to the ethical and fairness issues, experiments conducted in France and Europe (Aguzzi 2020) have so far shown reasoning biases leading to aberrant results. The use of algorithms for predictions could flout the principle of sentences in criminal matters and reproduce or reinforce inequalities.

This raises other issues. Firstly, the operation of existing algorithms remains in the hands of their designers due to business secrecy (Singh et al. 2021). Transparency about their design is highly limited, even though justice actors have embryonic expertise to evaluate them. Secondly, because algorithms ignore the possible interpretation of texts and the social context of judgements, the use of AI in justice entails a risk of standardising legal decisions.

## 14.4 Errors and Risks of the Interference of Artificial Intelligence in Legal Procedures: Integrating the Gap Between the Calculated World and the Sensible World

Meunier et al. (2019) propose an innovative approach extending to algorithms the phenomena of malfunctions affecting electrical systems, machines, software and interfaces (Berti-Equille et al. 2011). The aim is to identify the “bias zones” into which legal algorithms can fall, due to either structural errors or faulty assumptions. The goal is to detect when the algorithm fails either by going beyond the expectations of its designer or deviating from or misinterpreting its underlying assumptions. As AI learns from its own actions, it only takes one faulty element for the whole process to become biased. This could generate false positives which have little impact when it comes to not authorising a payment, but which, when applied to justice, can have otherwise serious effects.

With the progressive “datafication” of all human acts, data is at the heart of the problem. Are the algorithmic biases that often appear during the implementation of this type of process linked to the data collected, to the nature of the data collected, or to the methodologies for using the data? In other words, is it a question of the quality of the data, of the use of the data, or both?

This question was not born with AI. This goes back to the early days of cybernetics—the science of human government—which led people to believe that a society ‘steered and managed by machines’ would help humans avoid repeating the futile and bloody conflicts of history (Bateson 1972). This assumption ignored that behind the machines there were always humans, and they could only structure the machines according to their own conditioning or culture, reproducing or even reinforcing existing societal inequalities. Thus, many of the indicators that underpin medical decision-making and yet are used every day are—despite themselves—racialised (Cerdeña et al. 2020). There is no reason why justice and its automations should escape this problem (O’Neil 2016). AI gives greater resonance to these biases: it sorts and directs the choices of social actors, including judges, helps them make decisions, assigns indicators of recidivism, and supports recruitment processes. However, the points of vigilance that should be present at all levels of data collection, construction and interpretation are absent. Only the method of calculation and its efficiency or non-efficiency are addressed.

However, O’Neil, like other authors (Richard 2018), has faced much criticism, on the grounds that new algorithms would smooth out old biases until a form of quasi-automated ‘algorithmic fairness’ was achieved, with differences that would be measurable using dedicated indicators (Chadli et al. 2021). In other words, to improve technology, more technology is needed.

The question of bias is thus answered by calculated answers, even though the bodies of data that feed these systems remain imbued with pre-existing structural inequalities. Thus, algorithmic systems are trained on data that correspond to categorisations that have been the subject of very human choices. AI has not created labels like “individual at high risk of recidivism” or “having a cardiovascular risk”.

Ensuring a neutral reading of data does not guarantee its quality or accuracy. If the algorithm notices a bias in a body of data, it should refrain from using this source or correct the biased features, i.e. be able to look critically at its data source. Although data research has taken up the subject (Chereni et al. 2020), tools using AI for decision-making purposes are already being deployed on a large scale. Predictive policing, too, works with “manipulated” data because it relies on information about decades of past convictions without considering that these themselves may be the result of past societal biases (Noriega 2020). The same scenario can be found for algorithms used in the justice sector to detect the risk of recidivism (and thus possibly increase sentences). These are known to be largely unfavourable to certain populations. Even when smoothing out the first instance biases, Cyclic Redundant Coding (CRC) practices dedicated to the detection of errors in databases make it possible to recover traits by verifying other typologies of (socio-demographic) indicators that illustrate membership of given groups in the population. Information can be unintentionally modified during its transmission or storage in memory. Codes must therefore be used to detect or even correct errors due to these modifications. These codes cover more information than that strictly necessary to encode the information. To ‘m’ data bits, we add ‘k’ control bits. This means that  $n = m + k$  bits will be transmitted or stored in memory. These are known as redundant codes (Ntoutsu et al. 2020). Currently, some methodologies exist to support choices that are free from influences linked to the socio-historical context and economic imperatives. Indeed, even before tackling the technical part, ethical issues, such as the protection of people and their rights, must be integrated into the design of AI processes, as this will influence their architecture and the so-called “learning model” (supervised or unsupervised). In a supervised model, the training data is mastered by the designers who statistically structured the concerned dataset (this is called “data prep”) to avoid discrepancies (adequate population sizes and distributions). In an unsupervised model, the neural network itself detects bias. The network must learn to identify what is or can be a bias. The second important focus must be on the training dataset, the information used to teach the network how to behave. Usually, this is part of the global dataset used. If this dataset is biased (statistically not sound), the training will suffer the same failures. As such, a dataset including certain labels (race consideration for example) should not be used. This operation must be done with very special care at the three levels of data processing (pre -, in- and post), and all data that might introduce bias should be excluded in this very moment. Finally, a neural network is never fully trained. Periodic checks for bias in the results are important. AI tools can lose efficiency and performance over time, which, in the case of automated court decisions, could have serious societal impacts. Finally, and often forgotten, AI infrastructure is just as fundamental as the data itself. Collections made in real-time (e.g. stock market prices) can suffer from delays in their injection time into the processes, computers can break down, and, in the end, each grain of sand can bias the outputs.

The potential fallibility of AI’s predictability is precisely the reason for creating points of vigilance. The choices that prejudge the design of the tools deployed today actually raise questions that are more ethical than technical. These considerations can be reconsidered by abandoning the idea that society would be powerless in the face

of AI's opaque functioning. The tools currently deployed in the field of justice are far from neutral. To ensure the implementation of balanced or neutralised algorithms, three levels of intervention must be considered:

- The data
- The data integrity test
- The calculation, i.e. the algorithm.

These three layers are well known and understood on an individual basis by deep learning domain specialists. We know that ethical AI can be deployed; the only obstacle is the societal will to do so.

## 14.5 Recommendations by Way of Conclusion

In applying AI to the judicial and law enforcement domains, the question of the promoters of these processes arises. In concrete terms, this amounts to questioning the decision-making power in the field, both in terms of the structure and the reasons inherent to deploying this type of tool.

In law and justice, most people point to their multiple advantages. LegalTech advocates keep repeating that by automating certain tasks, AI would lighten the daily lives of judges and lawyers by relieving them of certain administrative constraints. Judges could also rely on AI for decision support, while lawyers would refine their strategies thanks to the case analyses proposed by AI. As for litigants, the creation of databases from court decisions would give them access to a real legal library. They could see the cost of their proceedings fall due to simplified management, and the time taken to appear in court would decrease.

The problem is that while the costs of implementation are known, the benefits, both operational and financial, for stakeholders other than the promoters of these tools remain unclear (Tung 2019). For example, although the potential financial savings from these deployments are frequently discussed (Rigano 2019), statistics on the reality of these gains are sorely lacking. The cost/benefit trade-off should be reviewed in the light of transparent economic impact studies, as it is not certain that systematisation in this area would be genuinely efficient at optimising the resource use, notably public judicial funds. Regardless, if it seems that AI would improve support for the legal process, it is unclear whether it can be considered a tool for justice in the full sense of the word at present (Wachter et al. 2021).

Certainly, AI could have macroeconomic effects (Mateu and Pluchart 2019). Political and institutional support for LegalTech actors as a well-thought-out tool for international leadership can be agreed upon. This would be a continuation of the multi-secular struggle of legal systems around the world (Ogus 1999). But questions remain as to whether these reasons are of sufficient importance to the general interest to justify and continue supporting technical arbitrations that are known to be biased (Pasquale 2020).



By combining the recommendations of the EU strategy for AI (Act, A. I. EUR-Lex-52021PC0206, 2021), especially point 3.5 on fundamental rights, the institutional studies carried out on this subject (Spajosevic et al. 2020), the analyses of the Council of Europe (CEPEJ 2020b), and the visions of major specialists in the field (Pasquale 2020; O’Neil 2016), certain features of a broad response to the problems mentioned above are emerging. They can be summarised in four points, all of which are based on a logic of ‘guaranteed transparency’ of systems.

***A—Knowledge of the production triptych: the promoter/financier, the designer, the developer***

An AI is only a tool that will answer a given question. The more societal the question, the more neutral the AI must be in its learning to be accepted. This implies moderation operations that have yet to be deployed, specifically to avoid what is called “false positives” As a reminder, a false positive is the Bayesian result of making a decision in a two-choice situation that is declared positive when it is actually negative. The outcome may be the result of a hypothesis test, an automatic classification algorithm, or even an arbitrary choice. However, if false positives are tolerable in terms of payment by bank card (after all, it can prevent compulsive purchases), it is unacceptable in terms of justice because it attacks the very foundations of equality, and therefore the balance of our societies. Destroying the balance amounts to destroying all confidence in social systems that are already under attack from populists.

Each AI deployed in the legal field, and therefore likely to become a tool for generating evidence, should have its promoters and developers identified so that the results of this type of activity can be integrated into the adversarial process in the same way as other types of evidence. Without falling into any kind of naivety, calculated evidence should be challenged in the same way as observed or recorded evidence.

***B—Knowledge of the essential characteristics of the algorithms***

The nature of information processing (supervised or unsupervised learning), and of the associated processes of moderation, whether human or calculated (so-called ‘neutralising’ bias), should be made public if only to demonstrate their existence.

***C—Knowledge of the sources, nature and architecture of the data used to train the algorithms***

Whether data is structured or unstructured, and in particular the so-called data-prep processes, should be made clear. This would make it possible to test the quality of the data and therefore the reliability of the foundations of the AI.

***D—Knowledge of uses to avoid unanticipated uses***

Any technology can be misused. It is not a question, as some suggest, of issuing authorisations for use, but of knowing who uses this type of tool and why.

All this information could be deposited in a supra-national registry (perhaps carried out by the EU) in return for which applicants would be granted rights or certificates of use as suggested by the CEPEJ in its feasibility study (CEPEJ 2020c). Once this knowledge ecosystem has been established, the possible deployment and



further uses of an algorithmic approach to justice can be imagined, ensuring that artificial intelligence is put to the service of all.

## References

- Act, A.I. 2021. Proposal for a regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. *EUR-Lex-52021PC0206*. <https://eur-lex.europa.eu/legal-content/EN/HIS/?uri=CELEX:52021PC0206>. Accessed 19 December 2021.
- Alston, P. 2019. Digital technology, social protection and human rights: Report. OHCHR. <https://www.ohchr.org/EN/Issues/Poverty/Pages/DigitalTechnology.aspx>. Accessed 17 May 2021.
- Alsop, T., S. Calio, and P. Greis. 2020. LEGAL TECH: A Statista dossier on the worldwide Legal Tech market. <https://www.statista.com/study/84775/legal-tech>. Accessed 4 December 2022.
- Aguzzi, C. 2020. Le juge et l'intelligence artificielle: La perspective d'une justice rendue par la machine. *Annuaire International De Justice Constitutionnelle* 35 (2019): 621–636.
- Barocas, S., and M. Hardt. 2017. NIPS tutorial on fairness in machine learning. <https://fairmlbook.org/tutorial1.html>. Accessed 4 December 2022.
- Bateson, G. 1972. From Versailles to cybernetics. In *Steps to an Ecology of Mind*, 477–485. London: Jason Aronson, Inc.
- Bathae, Y. 2017. The artificial intelligence black box and the failure of intent and causation. *Harvard Journal of Law & Technology* 31 (2): 889–938.
- Bell, B.S., A.M. Ryan, and D. Wiechmann. 2004. Justice expectations and applicant perceptions. *International Journal of Selection and Assessment* 12 (1–2): 24–38.
- Bennett, A. 2019. Explainer—Deanna Amato's robo-debt case. Victoria Legal Aid. <https://www.legalaid.vic.gov.au/about-us/news/explainer-deanna-amatos-robo-debt-case>. Accessed 17 May 2021.
- Berti-Equille, L., T. Dasu, and D. Srivastava. 2011. Discovery of complex glitch patterns: A novel approach to quantitative data cleaning. In *2011 IEEE 27th International Conference on Data Engineering*, 733–744. IEEE.
- Carney, T. 2020. Artificial intelligence in welfare: Striking the vulnerability balance? *Monash University Law Review* 46 (2): 23–51. <https://doi.org/10.26180/13370369.v2>.
- CEPEJ (Commission Européenne Pour l'Efficacité de la Justice). 2020a. European judicial systems - CEPEJ evaluation report - 2020a evaluation cycle (2018 data). <https://www.coe.int/fr/web/cepej/special-file-publication-of-the-report-european-judicial-systems-cepej-evaluation-report-2020a-evaluation-cycle-2018-data->. Accessed 1 December 2021.
- CEPEJ (Commission Européenne Pour l'Efficacité de la Justice). 2020b. CEPEJ tools on evaluation of judicial systems. <https://www.coe.int/en/web/cepej/eval-tools>. Accessed 25 November 2021.
- CEPEJ (Commission Européenne Pour l'Efficacité de la Justice). 2020c. Mise en place éventuelle d'un mécanisme de certification des outils et services d'intelligence artificielle dans le domaine juridique et judiciaire, étude de faisabilité réalisée par la Commission européenne pour l'efficacité de la justice, 8 décembre 2020c, CEPEJ(2020c)15 Rev. <https://rm.coe.int/etude-faisabilite-fr-cepej-2020-15/1680a0adf3>. Accessed 4 March 2021.
- Cerdeña, J.P., M.V. Plaisime, and J. Tsai. 2020. From race-based to race-conscious medicine: How anti-racist uprisings call us to act. *The Lancet* 396 (10257): 1125–1128.
- Chadli, S., P. Neveux, and T. Real. 2021. Intelligence Artificielle et éthique: comment définir et mesurer l'équité algorithmique? <https://www.quantmetry.com/blog/intelligence-artificielle-et-ethique-comment-definir-et-mesurer-lequite-algorithmique/>. Accessed 25 November 2021.
- Chereni, S., R.V. Sliuzas, and J. Flacke. 2020. An extended briefing and debriefing technique to enhance data quality in cross-national/language mixed-method research. *International Journal of Social Research Methodology* 23 (6): 661–675.

- Coulon, J.M. 2008. *La dépenalisation de la vie des affaires: rapport au garde des sceaux, ministre de la justice*. La Documentation française. [http://www.presse.justice.gouv.fr/art\\_pix/1\\_Rapport\\_Coulon.pdf](http://www.presse.justice.gouv.fr/art_pix/1_Rapport_Coulon.pdf). Accessed 4 December 2022.
- Croze, H. 2017. Justice prédictive: La factualisation du droit. *La semaine juridique-édition générale* 5: comm-101.
- Delsol, J-P. 2021. Quand l'état de droit se fissure, la démocratie frissonne. *Les Echos* 3 Septembre 2021. <https://www.lesechos.fr/idees-debats/cercle/opinion-quand-letat-de-droit-se-fisse-sure-la-democratie-frissonne-1343029>. Accessed 25 November 2021.
- Ellis, A. 2018. Neutrality and the civil service. In *Liberal Neutrality*, ed. Andrew Reeve and Robert E. Goodin, 92–113. New York: Routledge.
- European Commission. 2021. The 2021 EU Justice Scoreboard. Communication from the Commission to the European Parliament, the Council, the European Central Bank, the European Economic and Social Committee and the Committee of the Regions COM(2021) 389. [https://ec.europa.eu/info/sites/default/files/eu\\_justice\\_scoreboard\\_2021.pdf](https://ec.europa.eu/info/sites/default/files/eu_justice_scoreboard_2021.pdf). Accessed 18 October 2021.
- Fallery, B. 2019. Intelligence Artificielle: à qui profitent les ambiguïtés entre l'informatique, l'humain et la fiction? In *Journées IP&M sur L'Intelligence artificielle*. <https://hal.archives-ouvertes.fr/hal-03126061>. Accessed 31 December 2021.
- Ferrand, F. 2000. Le principe contradictoire et l'expertise en droit comparé européen. *Revue Internationale De Droit Comparé* 52 (2): 345–369.
- Ferrand, F. 2015. L'offre de médiation en Europe : Morceaux choisis. *Revue Internationale De Droit Comparé* 67 (1): 45–84.
- Gray, D. 2008. The people's courts? Summary justice and social relations in the city of London, c. 1760–1800. *Family & Community History* 11(1): 7–15.
- Guggeis, M. 2020. The Responsibility of the European Legislator and of the National Parliaments for Improving Relations between European Citizens and EU Institutions. In *The crisis of confidence in legislation*, ed. Maria De Benedetto, Nicola Lupo, and Nicoletta Rangone, 259–284. Baden Baden: Nomos Verlagsgesellschaft mbH & Co. KG.
- Hillebrandt, M., and S. Novak. 2016. 'Integration without transparency'? Reliance on the space to think in the European Council and Council. *Journal of European Integration* 38 (5): 527–540.
- Hyde, A. 2019. Vers une cyber-éthique de la justice dite prédictive» Commentaire de la Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement de la CEPEJ du 4 décembre 2018. *Daloz IP/IT*. <https://hal.archives-ouvertes.fr/hal-02395659>. Accessed 31 December 2021.
- Jaafari, M. 2016. Depenalization of business law. *The Judiciary Law Journal* 80 (94): 95–116.
- Keyvanpour, M.R., M. Javideh, and M.R. Ebrahimi. 2011. Detecting and investigating crime by means of data mining: A general crime matching framework. *Procedia Computer Science* 3: 872–880.
- Katz, P. S. 2014. Expert robot: Using artificial intelligence to assist judges in admitting scientific expert testimony. *Albany Law Journal of Science and Technology* 24(1).
- Langford, M. 2020. Taming the digital leviathan: automated decision-making and international human rights. *AJIL Unbound* 114: 141–146. <https://doi.org/10.1017/aju.2020.31>.
- Mateu, J., and J. Pluchart. 2019. L'économie de l'intelligence artificielle. *Revue D'économie Financière* 135: 257–272. <https://doi.org/10.3917/ecofi.135.0257>.
- Maurel, L. 2019. Filtrage automatique et libertés: peut-on sortir d'un Internet centralisé? *Annales des Mines-Enjeux Numériques* 6.
- Meunier, Axel, Donato Ricci, Dominique Cardon, and Maxime Crépel. 2019. Les glitches, ces moments où les algorithmes tremblent. *Technique & Culture* 72: 200–203. <https://doi.org/10.4000/tc.12698>.
- Noriega, M. 2020. The application of artificial intelligence in police interrogations: An analysis addressing the proposed effect AI has on racial and gender bias, cooperation, and false confessions. *Futures* 117: 102510.
- Ntoutsis, E., P. Fafalios, U. Gadiraju, V. Iosifidis, W. Nejdil, M.E. VidalS, F. Ruggieri, S. Turini, E. Papadopoulos, I. Krasanakis, K. Kompatsiaris, C. Kinder-Kurlanda, F. Wagner, M. Karimi, H.

- Fernandez, B. Alani, T. Berendt, C. Kruegel, K. Heinze, G. Broelemann, T. Tiropanis, Kasneci, and S. Staab. 2020. Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10 (3): e1356.
- Ogus, A. 1999. Competition between national legal systems: A contribution of economic analysis to comparative law. *International & Comparative Law Quarterly* 48 (2): 405–418.
- O’Neil, C. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown.
- Parviainen, J., A. Koski, and S. Torkkola. 2021. ‘Building a Ship while Sailing It.’ Epistemic Humility and the Temporality of Non-knowledge in Political Decision-making on COVID-19. *Social Epistemology* 35(3): 232–244.
- Pasquale, F. 2019. A rule of persons, not machines: the limits of legal automation. *George Washington Law Review* 87(1).
- Pasquale, F. 2020. *New Laws of Robotics*. Cambridge, Massachusetts: Harvard University Press.
- Poizat, J.C. 2017. Nouvelles réflexions sur la «banalité du mal». Autour du livre de Hannah Arendt Eichmann à Jérusalem et de quelques malentendus persistants à son sujet. *Le Philosophoire* 2: 233–252.
- Prakken, H., and G. Sartor. 2002. The role of logic in computational models of legal argument: a critical survey. *Computational logic: Logic programming and beyond* 342–381.
- Queudot, M., and M. J. Meurs. 2018. Artificial intelligence and predictive justice: Limitations and perspectives. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, eds. M. Mouhoub, S. Sadaoui, O. Ait Mohamed, and M. Ali, 889–897. Cham: Springer. doi: [https://doi.org/10.1007/978-3-319-92058-0\\_85](https://doi.org/10.1007/978-3-319-92058-0_85).
- Richard, C. 2018. Dans la boîte noire des algorithmes. *Revue Du Crieur* 3: 68–85.
- Rigano, C. 2019. Using artificial intelligence to address criminal justice needs. *National Institute of Justice*. <https://nij.ojp.gov/topics/articles/using-artificial-intelligence-address-criminal-justice-needs>. Accessed 31 December 2021.
- Ross, T. 1933. Machines that think. *Scientific American* 148 (4): 206–208.
- Singh, Y., P.K. Bhatia, and O. Sangwan. 2007. A review of studies on machine learning techniques. *International Journal of Computer Science and Security* 1 (1): 70–84.
- Singh, N., T.K. Bandyopadhyay, N. Sahoo, and K. Tiwari. 2021. Intellectual property issues in artificial intelligence: Specific reference to the service sector. *International Journal of Technological Learning, Innovation and Development* 13 (1): 82–100.
- Spajosevic, D., A. Ittoo, L. Rebouh, and E. de Kerchove. 2020. *Study on the use of innovative technologies in the justice field*. European Commission. <https://doi.org/10.2838/585101>. Accessed 31 December 2021.
- Stranieri, A., J. Zeleznikow, M. Gawler, and B. Lewis. 1999. A hybrid rule–neural approach for the automation of legal reasoning in the discretionary domain of family law in Australia. *Artificial Intelligence and Law* 7 (2): 153–183.
- Stoett, P.J. 2019. *Global ecopolitics: Crisis, governance, and justice*. Toronto: University of Toronto Press.
- T-540/15 Decision ECLI:EU:T:2018:167 22/03/2018 De Capitani/Parliament. <https://curia.europa.eu/juris/document/document.jsf?text=space%2Bto%2Bthink&docid=200551&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=2072856#ctx1>. Accessed 7 February 2021.
- Thagard, P. 2004. Causal inference in legal decision making: Explanatory coherence vs. *Bayesian Networks*. *Applied Artificial Intelligence* 18 (3–4): 231–249.
- Towell, N. 2017. Centrelink’s robo-debt crisis: Who is Hank Jongen? *Canberra Times*. <https://www.canberratimes.com.au/story/6037265/centrelinks-robo-debt-crisis-who-is-hank-jongen/>. Accessed 10 November 2020.
- Tung, K. 2019. AI, the internet of legal things, and lawyers. *Journal of Management Analytics* 6 (4): 390–403.
- Ulenaers, J. 2020. The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge? *Asian Journal of Law and Economics* 11(2).

Wachter, S., B. Mittelstadt, and C. Russell. 2021. Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review* 41: 105567.

Whittlestone, J., R. Nyrup, A. Alexandrova, K. Dihal, and S. Cave. 2019. Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. *Nuffield Foundation*. <https://www.nuffieldfoundation.org/about/publications/ethical-and-societal-implications-of-algorithms-data-and-artificial-intelligence-a-roadmap-for-research>. Accessed 31 December 2021.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

