Law

The law affects and is affected by the development and deployment of autonomous and intelligent systems (A/IS) in contemporary life. Science, technological development, law, public policy, and ethics are not independent fields of activity that occasionally overlap. Instead, they are disciplines that are fundamentally tied to each other and collectively interact in the creation of a social order.

Accordingly, in studying A/IS and the law, we focus not only on how the law responds to the technological innovation represented by A/IS, but also on how the law guides and sets the conditions for that innovation. This interactive process is complex, and its desired outcomes can rest on particular legal and cultural traditions. While acknowledging this complexity and uncertainty, as well as the acute risk that A/IS may intentionally or unintentionally be misused or abused, we seek to identify principles that will steer this interactive process in a manner that leads to the improvement, prosperity, and well-being of everyone.

The fact that the law has a unique role to play in achieving this outcome is observed by Sheila Jasanoff, a preeminent scholar of science and technology studies:

Part of the answer is to recognize that science and technology—for all their power to create, preserve, and destroy—are not the only engines of innovation in the world. Other social institutions also innovate, and they may play an invaluable part in realigning the aims of science and technology with those of culturally disparate human societies. Foremost among these is the law.[1]

The law can play its part in ensuring that A/IS, in both design and operation, are aligned with principles of ethics and human well-being.[2]

Comprehensive coverage of all issues within our scope of study is not feasible in a single chapter of *Ethically Aligned Design (EAD)*. Accordingly, aggregate coverage will expand as issues not yet studied are selected for treatment in future versions of *EAD*.

Law

EAD, First Edition includes commentary about how the law should respond to a number of specific ethical and legal challenges raised by the development and deployment of A/IS in contemporary life. It also focuses on the impact of A/IS *on the practice of law itself*. More specifically, we study both the potential benefits and the potential risks resulting from the incorporation of A/IS into a society's legal system—specifically, in law making, civil justice, criminal justice, and law enforcement. Considering the results of those inquiries, we endeavor to identify norms for the adoption of A/IS in a legal system that will enable the realization of the benefits while mitigating the risks.[3]

In this chapter of EAD, we include the following:

Section 1: Norms for the Trustworthy Adoption of A/IS in Legal Systems.
This section addresses issues raised by the potential adoption of A/IS in legal systems for the purpose of performing, or assisting in performing, tasks traditionally carried out by humans with specialized legal training or expertise. The section begins with the question of how A/IS, if properly incorporated into a legal system, can improve the functions of that legal system and thus enhance its ability to contribute to human well-being. The section then discusses challenges to the safe and effective incorporation of A/IS into a legal system and identifies **the chief challenge as an absence of informed trust.** The remainder of the section examines how societies can fill the trust gap by enacting policies and promoting practices that advance publicly accessible standards of **effectiveness**, **competence**, **accountability**, and **transparency**.

Section 2: Legal Status of A/IS.
This section addresses issues raised by the legal status of A/IS, including the potential assignment of certain legal rights and obligations to such systems. The section provides background on the issue and outlines some of the potential advantages and disadvantages of assigning some form of legal personhood to A/IS. Based on these considerations, the section concludes that extending legal personhood to A/IS is not appropriate at this time. It then considers alternatives and outlines certain future conditions that might warrant reconsideration of the section's central recommendation.

Law

Section 1: Norms for the Trustworthy Adoption of A/IS in Legal Systems[4]

"It's a day that is here."
John G. Roberts , Chief Justice of the Supreme Court of the United States, when asked in 2017 whether he could foresee a day when intelligent machines would assist with courtroom fact-finding or judicial decision-making.[5]

A/IS hold the potential to improve the functioning of a legal system and, thereby, to contribute to human well-being. That potential will be realized, however, only if both the use of A/IS and the avoidance of their use are grounded in solid information about the capabilities and limitations of A/IS, the competencies and conditions required for their safe and effective operation (including data requirements), and the lines along which responsibility for the outcomes generated by A/IS can be assigned. Absent that information, society risks both **uninformed adoption** of A/IS and **uninformed avoidance of adoption** of A/IS, risks that are particularly acute when A/IS are applied in an integral component of the social order, such as the law.

- **Uninformed adoption** poses the risk that A/IS will be applied to inform or replace the judgments of legal actors (legislators, judges, lawyers, law enforcement officers, and jurors) without controls to ensure their safe and effective operation. They may even be used for purposes other than those for which the systems have been validated and vetted for legal use. In addition to actual harm to individuals, the result will be distrust, not only of the effectiveness of A/IS, but also of the fairness and effectiveness of the legal system itself.
- **Uninformed avoidance of adoption** poses the risk that a lack of understanding of what is required for the safe and effective operation of A/IS will result in blanket distrust of all forms and applications of A/IS, even those that are, when properly applied, safe and effective. The result will be a failure to realize the significant improvements in the legal system that A/IS can offer and a continuation of systems that are, even with the best of safeguards, still subject to human bias, inconsistency, and error.[6]

In this section, we consider how society can address these risks by developing norms for the adoption of A/IS in legal systems. The specific issues discussed follow. The first and second issues reflect the potential benefits of, and challenges to, trustworthy adoption of A/IS in the world's legal systems. The remaining issues discuss four principles,[7] which, if adhered to, will enable trustworthy adoption.[8] [9]

Law

- **Issue 1: Well-being, Legal Systems, and A/IS—**How can A/IS improve the functioning of a legal system and, thereby, enhance human well-being?
- **Issue 2: Impediments to Informed Trust—**What are the challenges to adopting A/IS in legal systems and how can those impediments be overcome?
- **Issue 3: Effectiveness—**How can the collection and disclosure of evidence of effectiveness of A/IS foster informed trust in the suitability of A/IS for adoption in legal systems?
- **Issue 4: Competence—**How can specification of the knowledge and skills required of the human operator(s) of A/IS foster informed trust in the suitability of A/IS for adoption in legal systems?
- **Issue 5: Accountability—**How can the ability to apportion responsibility for the outcome of the application of A/IS foster informed trust in the suitability of A/IS for adoption in legal systems?
- **Issue 6: Transparency—**How can sharing information that explains how A/IS reach given decisions or outcomes foster informed trust in the suitability of A/IS for adoption in legal systems?

Issue 1: Well-Being, Legal Systems, and A/IS

How can A/IS improve the functioning of a legal system and, thereby, enhance human well-being?

Background

An effective legal system contributes to human well-being. The law is an integral component of social order; the nature of a legal system informs, in fundamental ways, the nature of a society, its potential for economic growth and technological innovation, and its capacity for advancing the well-being of its members.

If the law is a constitutive element of social order, it is not surprising that it also plays a key role in setting the conditions for well-being and economic growth. In part, this flows from the fact that a well-functioning legal system is an element of good governance. Good governance and a well-functioning legal system can help society and its members flourish, as measured by indicators of both economic prosperity[10] and human well-being.[11] The attributes of good governance can be defined in several ways. Good governance can mean democracy; the observance of norms of human rights enshrined in conventions such as the Universal Declaration of Human Rights[12] and the Convention of the Rights of the Child;[13] and constitutional constraints on government power. It can also

Law

mean bureaucratic competence, law and order, property rights, and contract enforcement.

The United Nations (UN) defines the rule of law as:

> a principle of governance in which all persons, institutions and entities, public and private, including the State itself, are accountable to laws that are publicly promulgated, equally enforced and independently adjudicated. . . . It requires, as well, measures to ensure adherence to the principles of supremacy of law, equality before the law, accountability to the law, fairness in the application of the law, separation of powers, participation in decision-making, legal certainty, avoidance of arbitrariness and procedural and legal transparency.[14]

Orderly systems of legal rules and institutions generally correlate positively with economic prosperity, social stability, and human well-being, including the protection of childhood.[15] Studies from the World Bank suggest that legal reforms can lead to increased foreign investment, higher incomes, and greater wealth.[16] Wealth, in turn, can enable policies that support improved education, health, environmental protection, equal opportunity, and, in democratic societies, greater individual freedom.

Law, moreover, can contribute to prosperity not only through its functional attributes, but also through its substantive content. Patent laws, for example, if well-designed, can encourage technological innovation, leading to increases in productivity and the economic growth that follows. Poorly designed patent laws, on the other hand, may foster monopolistic markets and decrease competition, resulting in a decreased pace of technological innovation, fewer gains in productivity, and slower economic growth.[17]

While economic growth is a valuable benefit of a well-designed and well-functioning legal system, it is not the only benefit. Such a system can bring benefits to society and its members that, beyond economic prosperity, extend to mental and physical well-being. Specific benefits include the protection and advancement of an individual's dignity,[18] human rights,[19] liberty, stability, security, equality of treatment under the law, and ability to provide for the future.[20]

In fact, recent thinking on the relationship between law and economic development has come to hold that a well-functioning legal system is not simply a *means* to development but *is* development, insofar as such a system is a constitutive element of a social order that protects and advances human dignity, rights, and well-being. As this position has been characterized by David Kennedy:

> … the focal point for development policy was increasingly provided less by economics than from ideas about the nature of the good state themselves provided by literatures of political science, political economy, ethics, social theory, and law. In particular, "human rights" and the "rule of law"[21] became substantive definitions of development. One should promote human rights not to *facilitate* development—but *as* development. The rule of law was not a development *tool*—it was itself a development

Law

objective. Increasingly, law—understood as a combination of human rights, courts, property rights, formalization of entitlements, prosecution of corruption, and public order—came to define development.[22]

While this shift from considering law as a means to an end to considering law as an end in itself has been criticized on the grounds that it takes the focus off the difficult political choices that are inherent in any development policy,[23] it remains true that a well-functioning legal system is essential to the realization of a social order that protects and advances human dignity, rights, and well-being.

A/IS can contribute to the proper functioning of a legal system. A properly functioning legal system, one that is conducive to both economic prosperity and human well-being, will have a number of attributes. It should be:

- **Speedy:** enable quick resolution of civil and criminal cases;
- **Fair:** produce results that are just and proportionate to circumstance;[24]
- **Free from undesirable bias:** operate without prejudice;
- **Consistent:** arrive at outcomes in a principled, consistent, and nonarbitrary manner;
- **Transparent:** be open to appropriate public examination and oversight;[25]
- **Accessible:** be equally open to all citizens and residents in resolving disputes;
- **Effective:** achieve the ends intended by its laws and rules without negative collateral consequences;[26]
- **Accurate:** achieve accurate results, minimizing both false positives (persons unjustly or incorrectly targeted, investigated, or sentenced for crimes) and false negatives (persons incorrectly *not* targeted, investigated, or sentenced for crimes);
- **Adaptable:** have the flexibility to adapt to changes in societal circumstances.

A/IS have the potential to alter the overall functioning of a legal system. A/IS, applied responsibly and appropriately, could improve the legislative process, enhance access to justice, accelerate judicial decision-making, provide transparent and readily accessible information on why and how decisions were reached, reduce bias, support uniformity in judicial outcomes, help society identify (and potentially correct) judicial errors, and improve public confidence in the legal system. By way of example:

- A/IS can make legislation and regulation more **effective** and **adaptable**. For lawmaking, A/IS could help legislators analyze data to craft more finely tuned, responsive, evidence-based laws and regulations. This could, potentially, offer self-correcting suggestions to legislators (and to the general public) to help inform dialogue on how to meet defined public policy objectives.
- A/IS can make the practice of law more effective and efficient. For example, A/IS can enhance the **speed**, **accuracy**, and **accessibility** of the process of fact-finding in legal proceedings. When used appropriately in legal fact-finding, particularly in jurisdictions that allow extensive discovery or disclosure, A/IS already make litigation and investigations more accessible by analyzing vast data

Law

collections faster, more efficiently, and potentially more effectively[27] than document analysis conducted solely by human attorneys. By making fact-finding in an era of big data progressively easier, faster, and cheaper, A/IS may facilitate access to justice for parties that otherwise may find using the legal system to resolve disputes cost-prohibitive. A/IS can also help ensure that justice is rendered based on better accounting of the facts, thus serving the central purpose of any legal system.

- In both civil and criminal proceedings, A/IS can be used to improve the **accuracy, fairness,** and **consistency** of decisions rendered during proceedings. A/IS could serve as an auditing function for both the civil and criminal justice systems, helping to identify and correct judicial and law enforcement errors.[28]
- A/IS can increase the **speed**, **accuracy**, **fairness**, **freedom from bias**, and general **effectiveness** with which law enforcement resources are deployed to combat crime. A/IS could be used to reduce or prevent crime, respond more quickly to crimes in progress, and improve collaboration among different law enforcement agencies.[29]
- A/IS can help ensure that determinations about the arrest, detention, and incarceration of individuals suspected of, or convicted of, violations of the law are **fair**, **free from bias**, **consistent**, and **accurate**. Automated risk assessment tools have the potential to address issues of systemic racial bias in sentencing, parole, and bail determination while also safely reducing incarceration and recidivism rates by identifying individuals who are less likely to commit crimes if released.
- A/IS can help to ensure that the tools, procedures, and resources of the legal system are more **transparent** and **accessible** to citizens. For the ordinary citizen, A/IS can democratize access to legal expertise, especially in smaller matters, where they may provide effective, prompt, and low-cost initial guidance to an aggrieved party; for example, in landlord-tenant, product purchase, employment, or other contractual contexts where the individual often tends to find access to legal information and legal advice prohibitive, or where asymmetry of resources between the parties renders recourse to the legal system inequitable.[30]

A/IS have the potential to improve how a legal system functions in fundamental ways. As is the case with all powerful tools, there are some risks. **A/IS should not be adopted in a legal system without due care and scrutiny;** they should be adopted after a society's careful reflection and proper examination of evidence that their deployment and operation can be trusted to advance human dignity, rights, and well-being (see Issues 2–6).

Recommendations[31]

1. Policymakers should, in the interest of improving the function of their legal systems and bringing about improvements to human well-being, explore, through a broad consultative dialogue with all stakeholders, how A/IS can be adopted for use in their legal systems. They should do

so, however, only in accordance with norms for adoption that mitigate the risks attendant on such adoption (see Issues 2–6 in this section).

2. Governments, non-governmental organizations, and professional associations should support educational initiatives designed to create greater awareness among all stakeholders of the potential benefits and risks of adopting A/IS in the legal system, and of the ways of mitigating such risks. A particular focus of these initiatives should be the ordinary citizen who interacts with the legal system as a victim or criminal defendant.

Further Resources

- A. Brunetti, G. Kisunko, and B. Weder, "Credibility of Rules and Economic Growth: Evidence from a Worldwide Survey of the Private Sector," The World Bank Economic Review, vol. 12, no. 3, pp. 353-384, Sep. 1998.
- S. Jasanoff, "Governing Innovation: The Social Contract and the Democratic Imagination," Seminar, vol. 597, pp. 16-25, May 2009.
- D. Kennedy, "The 'Rule of Law,' Political Choices and Development Common Sense," in *The New Law and Economic Development: A Critical Appraisal*, D. M. Trubek and A. Santos, eds., Cambridge: Cambridge University Press, 2006, pp. 95-173.
- "Artificial Intelligence," National Institute of Standards and Technology.
- K. Schwab, "The Global Competitiveness Report: 2018," The World Economic Forum, 2018.
- A. Sen, Development as Freedom. New York, NY: Alfred A. Knopf, 1999.
- United Nations General Assembly, Universal Declaration of Human Rights, Dec. 10, 1948.
- UNICEF, Convention on the Rights of the Child, Nov. 4, 2014.
- United Nations Office of the High Commissioner: Human Rights, The Vienna Declaration and Programme of Action, June 25, 1993.
- World Bank, World Development Report 2017: Governance and the Law, Jan. 2017.
- World Justice Project, Rule of Law Index, June 2018.

Law

Issue 2: Impediments to Informed Trust

What are the challenges to adopting A/IS in legal systems and how can those impediments be overcome?

Background

Although the benefits to be gained by adopting A/IS in legal systems are potentially numerous (see the discussion of Issue 1), there are also significant risks that must be addressed in order for the A/IS to be adopted in a manner that will realize those benefits. The risks sometimes mirror expected benefits:

- the potential for opaque decision-making;
- the intentional or unintentional biases and abuses of power;
- the emergence of nontraditional bad actors;
- the perpetuation of inequality;
- the depletion of public trust in a legal system;
- the lack of human capital active in judicial systems to manage and operate A/IS;
- the sacrifice of the spirit of the law in order to achieve the expediency that the letter of the law allows;
- the unanticipated consequences of the surrender of human agency to nonethical agents;
- the loss of privacy and dignity;
- and the erosion of democratic institutions.[32]

By way of example:

- Currently, A/IS used in justice systems are not subject to uniform rules and norms and are often adopted piecemeal at the local or regional level, thereby creating a highly variable landscape of tools and adoption practices. Critics argue that, far from improving fact-finding in civil and criminal matters or eliminating bias in law enforcement, these tools have unproven accuracy, are error-prone, and may serve to entrench existing social inequalities. These tools' potential must be weighed against their pitfalls. These include unclear efficacy; incompetent operation; and potential impairment of a legal system's ability to adhere to principles of socioeconomic, racial, or religious equality, government transparency, and individual due process, to render justice in an informed, consistent, and fair manner.
- In the case of *State v. Loomis*, an important but not widely known case, the Wisconsin Supreme Court held that a trial court's use of an algorithmic risk assessment tool in sentencing did not violate the defendant's due process rights, despite the fact that the methodology used to obtain the automated assessment was not disclosed to either the court or the defendant.[33] A man received a lengthy sentence based in part on what an opaque algorithm thought of him. While the court considered many factors, and sought to balance competing societal values, this

Law

is just one case in a growing set of cases illustrating how criminal justice systems are being impacted by proprietary claims of trade secrets, opaque operation of A/IS, a lack of evidence of the effectiveness of A/IS, and a lack of norms for the adoption of A/IS in the extended legal system.

- More generally, humans tend to be subject to the cognitive bias known as "anchoring", which can be described as the excessive reliance on an initial piece of information. This may lead to the progressive, unwitting, and detrimental reliance of judges and legal practitioners on assessments produced by A/IS. This risk is compounded by the fact that A/IS are (and shall remain in the foreseeable future) nonethical agents, incapable of empathy, and thus at risk of being unable to produce decisions aligned with not just the letter of the law, but also the spirit of the law and reasonable regard for the circumstances of each defendant.
- The required technical and scientific knowledge to procure, deploy, and effectively operate A/IS, as well as that required to measure the ability of A/IS to achieve a given purpose without adverse collateral consequences, represent significant hurdles to the beneficial long-term adoption of A/IS in a legal system. This is especially the case when—as is the case presently—actors in the civil and criminal justice systems and in law enforcement may lack the requisite specialized technological or scientific expertise.[34]

Such risks must be addressed in order to ensure sustainable management and public oversight of what will foreseeably become an increasingly automated justice system.[35] The view expressed by the Organisation for Economic Co-operation and Development (OECD) in the domain of digital security that "robust strategies to [manage risk] are essential to establish the trust needed for economic and social activities to fully benefit from digital innovation"[36] applies equally to the adoption of A/IS in the world's legal systems.

Informed trust. If we are to realize the benefits of A/IS, we must trust that they are safe and effective. People board airplanes, take medicine, and allow their children on amusement park rides because they trust that the tools, methods, and people powering those technologies meet certain safety and effectiveness standards that reduce the risks to an acceptable level given the objectives and benefits to be achieved. This need for trust is especially important in the case of A/IS used in a legal system. The "black box" nature and lack of trust in A/IS deployed in the service of a legal system could quickly translate into a lack of trust in the legal system itself. This, in turn, may lead to an undermining of the social order. Therefore, if we are to improve the functioning of our legal systems through the adoption of A/IS, **we must enact policies and promote practices that allow those technologies to be adopted on the basis of informed trust**. Informed trust rests on a reasoned evaluation of clear and accurate information about the effectiveness of A/IS and the competence of their operators.[37]

Law

To formulate policies and standards of practice intended to foster informed trust, it is helpful, first, to identify principles applicable over the entire supply chain for the delivery of A/IS-enabled decisions and guidance, including design, development, procurement, deployment, operation, and validation of effectiveness, that, if adhered to, will foster trust. Once those general principles have been identified, specific policies and standards of practice can be formulated that encourage adherence to the principles in every aspect of a legal system, including lawmaking, civil and criminal justice, and law enforcement. Such principles, if they are to serve their intended purpose of informing effective policies and practices, must meet certain design criteria. Specifically, **the principles should be (a) individually necessary and collectively sufficient, (b) globally applicable but culturally flexible, and (c) capable of being operationalized in applicable functions of the legal system**. A set of principles that meets these criteria will provide an effective framework for the development of policies and practices that foster trust, while leaving considerable flexibility in the specific policies and standards of practice that a society chooses to implement in furthering adherence to the principles.

A set of four principles that we believe meets the design criteria just described are the following:

- **Effectiveness**: Adoption of A/IS in a legal system should be based on sound empirical evidence that they are fit for their intended purpose.
- **Competence**: A/IS should be adopted in a legal system only if their creators specify the skills and knowledge required for their effective operation and if their operators adhere to those competency requirements.
- **Accountability**: A/IS should be adopted in a legal system only if all those engaged in their design, development, procurement, deployment, operation, and validation of effectiveness maintain clear and transparent lines of responsibility for their outcomes and are open to inquiries as may be appropriate.
- **Transparency**: A/IS should be adopted in a legal system only if the stakeholders in the results of A/IS have access to pertinent and appropriate information about their design, development, procurement, deployment, operation, and validation of effectiveness.

In the remainder of Section 1, we elaborate on each of these principles. Before turning to a specific discussion of each, we add two further considerations that should be kept in mind when applying them collectively.

Differences in emphasis. While all four of the aforementioned principles will contribute to the fostering of trust, each principle will *not* contribute equally in every circumstance. For example, in many applications of A/IS, a well-established measure of effectiveness, obtained by proven and accepted methods, may go a considerable way to creating conditions for trust in the given application. In such a case, the other principles may add to trust, but they may not be necessary to establish trust. Or, to take another example, in some applications the role of the human operator may be minimal, while in other applications there will be extensive scope for

Law

human agency where competence has a greater role to play. In finding the right emphasis and balance among the four principles, policymakers and practitioners will have to consider the specific circumstances of A/IS.

Flexibility in implementation. It should be noted that we have addressed the four principles above at a rather high level and have not offered specific prescriptions of how adherence to the principles should be implemented. This is by design. Although adherence to all four principles is important, it is also important that, at the operational level, flexibility be allowed for the selection and implementation of policies and practices that (a) are in harmony with a given society's traditions, norms, and values; (b) conform with the laws and regulations operative in a given jurisdiction; and (c) are consistent with the ethical obligations of legal practitioners.

Recommendations

1. Governments should set procurement and contracting requirements that encourage parties seeking to use A/IS in the conduct of business with or for the government, particularly with or for the court system and law enforcement agencies, to adhere to the principles of effectiveness, competence, accountability, and transparency as described in this chapter. This can be achieved through legislation or administrative regulation. All government efforts in this regard should be transparent and open to public scrutiny.
2. Professionals engaged in the practice, interpretation, and enforcement of the law (such as lawyers, judges, and law enforcement officers), when engaging with or relying on providers of A/IS technology or services, should require, at a minimum, that those providers adhere to, and be able to demonstrate adherence to, the principles of effectiveness, competence, accountability, and transparency as described in this chapter. Likewise, those professionals, when operating A/IS themselves, should adhere to, and be able to demonstrate adherence to, the principles of effectiveness, competence, accountability, and transparency. Demonstrations of adherence to the requirements should be publicly accessible.
3. Regulators should permit insurers to issue professional liability and other insurance policies that consider whether the insured (either a provider or operator of A/IS in a legal system) adheres to the principles of effectiveness, competence, accountability, and transparency (as they are articulated in this chapter).

Further Resources

- "Criminal Law—Sentencing Guidelines—Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing—State v. Loomis, 881 N.W.2d 749 (Wis. 2016)," Harvard Law Review, vol. 130, no. 5, pp. 1530-1537, 2017.
- K. Freeman, "Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis," North Carolina Journal of Law and Technology, vol. 18, no. 5, pp. 75-76, 2016.

- "Managing Digital Security and Privacy Risk: Background Report for Ministerial Panel 3.2," Organisation for Economic Co-operation and Development (OECD) Directorate for Science, Technology, and Innovation: Committee on Digital Economy Policy, June 1, 2016.
- *State v Loomis*, 881 N.W.2d 749 (Wis. 2016), *cert. denied* (2017).
- "Global Governance of AI Roundtable: Summary Report 2018," World Government Summit, 2018.

Issue 3: Effectiveness

How can the collection and disclosure of evidence of effectiveness of A/IS foster informed trust in the suitability for adoption in legal systems?

Background

An essential component of trust in a technology is trust that it works and meets the purpose for which it is intended. We now turn to a discussion of the role that evidence of effectiveness, chiefly in the form of the results of a measurement exercise, can play in fostering informed trust in A/IS as applied in legal systems.[38] We begin with a general characterization of what we mean by *evidence of effectiveness*: what we are measuring, how we are measuring, what form our results take, and who the intended consumers of the evidence are. We then identify the specific features of the practice of measuring effectiveness that will enable it to contribute to informed trust in A/IS as applied in a legal system.

What constitutes evidence of effectiveness?

What we are measuring. In gathering evidence of effectiveness, we are seeking to gather empirical data that will tell us whether a given technology or its application will serve as an effective solution to the problem it is intended to address. Serving as an effective solution means more than meeting narrow specifications or requirements; it means that **the A/IS are capable of addressing their target problems in the real world**, which, in the case of A/IS applied in a legal system, are problems in the making, administration, adjudication, or enforcement of the law. It also means remaining practically feasible once collateral concerns and potential unintended consequences are taken into account.[39] To take a non-A/IS example, under the definition of effectiveness we are considering, for an herbicide to be considered effective, it must be shown not only to kill the target weeds, but also to do so without causing harm to nontarget plants, to the person applying the agent, and to the environment in general.

Under the definition above, assessing the effectiveness of A/IS in accomplishing the target task (narrowly defined) is not sufficient; it may also be necessary to assess the extent to which the A/IS are aligned with applicable

Law

laws, regulations, and standards,[40] and whether (and to what extent) they impinge on values such as privacy, fairness, or freedom from bias.[41] Whether such collateral concerns are salient will depend on the nature of the A/IS and on the particular circumstances in which they are to be applied.[42] However, it is only from such a complete view of the impact of A/IS that a balanced judgment can be made of the appropriateness of their adoption.[43]

Although the scope of an evaluation of effectiveness is broader than a narrowly focused verification that a specific requirement is met, it has its limits. There are measures of aspects of A/IS that one might find useful but that are outside the scope of effectiveness. For example, given frequently expressed concerns that A/IS will one day cross the limits of their intended purpose and overwhelm their creators and users, one might seek to define and obtain general measures of the autonomy of a system or of a system's capacity for artificial general intelligence (AGI). Although such measures could be useful—assuming they could be defined—they are beyond the scope of evaluations of effectiveness. Effectiveness is always tied to a target purpose, even if it includes consideration of the collateral effects of the manner of meeting that purpose.

What we are measuring is therefore a general "fitness for purpose".

How we measure. Evidence of effectiveness is typically gathered in one of two types of exercises:[44]

- **A single-system validation exercise** measures and reports on the effectiveness of a single system on a given task. In such an exercise, the system to be validated will typically have already carried out the target task on a given data set. The purpose of the validation is to provide empirical evidence of how successful the system has been in carrying out the task on that data set. Measurements are obtained by independent sampling and review of the data to which the system was applied. Once obtained, those metrics serve to corroborate or refute the hypothesis that the system operated as intended in the instance under consideration. An example of validation as applied to legal fact-finding would be a test of the effectiveness of A/IS that had been used to retrieve material relevant (as defined by the humans deploying the system) to a given legal inquiry from a collection of emails.
- **A multi-system (or benchmarking) evaluation** involves conducting a comparative study of the effectiveness of several systems designed to meet the same objective. Typically, in such a study, a test data set is identified, a task to be performed is defined (ideally, a task that models the real-world objectives and conditions for which the systems under evaluation have been designed[45]), the systems to be evaluated are used to carry out the task, and the success of each system in carrying out the task is measured and reported. An example of this sort of evaluation applied to a specific

real-world challenge in the justice system is the series of evaluations of the effectiveness of information retrieval systems in civil discovery, including A/IS, conducted as part of the US National Institute of Standards and Technology (NIST) Text REtrieval Conference (TREC) Legal Track initiative.[46]

The measurements obtained by both types of evaluation exercises are valuable. The results of a single-system validation exercise are typically more specific, answering the question of whether a system *was* effective in a specific instance. The results of a multi-system evaluation are typically more generic, answering the question of whether a system *can* be effective in real-world circumstances. Both questions are important, hence both types of evaluations are valuable.[47]

The form of results. The results of an evaluation typically take the form of a number—a quantitative gauge of effectiveness. This can be, for example, the decreased likelihood of developing a given medical condition; safety ratings for automobiles; recall measures for retrieving responsive documents; and so on. Certainly, qualitative considerations are not (and should not) be ignored; they often provide context crucial to interpreting the quantitative results.[48] Nevertheless, at the heart of the results of an evaluation exercise is a number, a metric that serves as a telling indicator of effectiveness.[49]

In some cases, the research community engaged in developing any new system will have reached consensus on salient effectiveness metrics. In other cases, the research community may not have reached a consensus, requiring further study. In the case of A/IS, given both their accelerating development and the fact that they are often applied to tasks for which the effectiveness of their human counterparts is seldom precisely gauged, we are often still at the stage of defining metrics. An example of an application of A/IS for which there is a general consensus around measures of effectiveness is legal electronic discovery,[50] where there is a working consensus around the use of the evaluation metrics referred to as "recall" and "precision".[51] Conversely, in the case of A/IS applied in support of sentencing decisions, a consensus on the operative effectiveness metrics does not yet exist.[52]

The consumers of the results. In defining metrics, it is important to keep in mind the consumers of the results of an evaluation of effectiveness. Broadly speaking, it is helpful to distinguish between two categories of stakeholders who will be interested in measurements of effectiveness:

- **Experts** are the researchers, designers, operators, and advanced users with appropriate scientific or professional credentials who have a technical understanding of the way in which a system works and are well-versed in evaluation methods and the results they generate.
- **Nonexperts** are the legislators, judges, lawyers, prosecutors, litigants, communities, victims, defendants, and system advocates whose work or legal outcomes may, even if only indirectly, be affected by the results

of a given system. These individuals, however, may not have a technical understanding of the way in which a system operates. Furthermore, they may have little experience in conducting scientific evaluations and interpreting their results.

Effectiveness metrics must meet the needs of *both* expert *and* nonexpert consumers.

- With respect to experts, the purpose of an effectiveness metric is *to advance both long-term research and more immediate product development, maintenance, and oversight*. To achieve that purpose, it is appropriate to define a fine-grained metric that may not be within the grasp of the nonexpert. Researchers and developers will be acting on the information provided by such a metric, so it should be tailored to their needs.
- With respect to nonexperts, including the general public, the purpose of an effectiveness metric is *to advance informed trust*, meaning trust that is based on sound evidence that the A/IS have met, or will meet, their intended objectives, taking into account both the immediate purpose and the contextual purpose of preserving and fostering important values such as human rights, dignity, and well-being. For this purpose, it will be necessary to define a metric that can serve as a readily understood summary measure of effectiveness. This metric must provide a simple, direct answer to the question of how effective a given system is. Automobile safety ratings are an example of this sort of metric. For automobile designers and engineers, the summary metrics are not sufficiently fine-grained to give immediately actionable information; for consumers, however, the metrics, insofar as they are accurate, empower them to make better-informed buying decisions.

For the purpose of fostering informed trust in A/IS adopted in the legal system, the most important goal is to establish a clear measure of effectiveness that can be understood by nonexperts. However, significant obstacles to achieving this goal include (a) developer incentives that prioritize research and development, along with the metrics that support such efforts, and (b) market forces that inhibit, or do not encourage, consumer-facing metrics. For those reasons, it is important that the selection and definition of the operative metrics draw on input not only from the A/IS creators but from other stakeholders as well; only under these conditions will a consensus form around the meaningfulness of the metrics.

What measurement practices foster informed trust?

By equipping both experts and nonexperts with accurate information regarding the capabilities and limitations of a given system, measurements of effectiveness can provide society with information needed to adopt and apply A/IS in a thoughtful, carefully considered, beneficial manner.[53]

In order for the practice of measuring effectiveness to realize its full potential for fostering trust and mitigating the risks of uninformed adoption and uninformed avoidance of adoption, it must have certain features:

Law

- **Meaningful metrics:** As noted above, an essential element of a measurement practice is a metric that provides an accurate and readily understood gauge of effectiveness. The metric should provide clear and actionable information as to the extent to which a given application has, or has not, met its objective so that potential users of the results of the application can respond accordingly. For example, in legal discovery, both recall and precision have done this well and have contributed to the acceptance of the use of A/IS for this purpose.[54]
- **Sound methods:** Measures of effectiveness must be obtained by scientifically sound methods. If, for example, measures are obtained by sampling, those sample-based estimates must be the result of sound statistical procedures that hold up to objective scrutiny.
- **Valid data:** Data on which evaluations of effectiveness are conducted should accurately represent the actual data to which the given A/IS would be applied and should be vetted for potential bias. Any data sets used for benchmarking or testing should be collected, maintained, and used in accordance with principles for the protection of individual privacy and agency.[55]
- **Awareness and consensus:** Measurement practices must not only be technically sound in terms of metrics, methods, and data, but they must also be widely understood and accepted as evidence of effectiveness.
- **Implementation:** Measurement practices must be both practically feasible and actually implemented, i.e., widely adopted by practitioners[56].
- **Transparency.** Measurement methods and results must be open to scrutiny by experts and the general public.[57] Without such scrutiny, the measurements will not be trusted and will be incapable of fulfilling their intended purpose.[58]

In seeking to advance informed trust in A/IS, policymakers should formulate policies and promote standards that encourage sound measurement practices, especially those that incorporate the key features.

Additional note. While in all circumstances all four principles discussed in this chapter (Effectiveness, Competence, Accountability, Transparency) will have something to contribute to the fostering of informed trust, it is not the case that in every circumstance all four principles will contribute equally to the fostering of trust. In some circumstances, a well-established measure of effectiveness, obtained by proven and accepted methods, may go a considerable way, on its own, in fostering trust in a given application—or distrust, if that is what the measurements indicate. In such circumstances, the challenges presented by the other principles, e.g., the challenge of adhering to the principle of transparency while respecting intellectual property considerations, may become of secondary importance.

Law

Illustration—Effectiveness

The search for factual evidence in large document collections in US civil or criminal proceedings has traditionally involved page-by-page manual review by attorneys. Starting in the 1990s, the proliferation of electronic data, such as email, rendered manual review prohibitively costly and time-consuming. By 2008, A/IS designed to substantially automate review of electronic data (a task known as "e-discovery") were available. Yet, adoption remained limited. Chief among the obstacles to adoption was a concern about the effectiveness, and hence defensibility in court, of A/IS in e-discovery. **Simply put, practitioners and courts needed a sound answer to a simple question: "Does it work?"**

Starting in 2006, the US NIST [59] conducted studies to assess that question.[60] The studies focused on, among others, two sound statistical metrics, both expressed as easy-to-understand percentages:[61,62]

- **Recall**, which is a gauge of the extent to which all the relevant documents were retrieved. For example, if there were 1,000 relevant documents to be found in the collection, and the review process identified 700 of them, then it achieved 70% recall.
- **Precision**, which is a gauge of the extent to which the documents identified as relevant by a process were actually relevant. For example, if for every two relevant documents the system captured, it also captured a nonrelevant one (i.e., a false positive), then it achieved 67% precision.

The studies provided empirical evidence that some systems could achieve high scores (80%) according to both metrics.[63] In a seminal follow-up study, Maura R. Grossman and Gordon V. Cormack found that two automated systems did, in fact, "conclusively" outperform human reviewers.[64] Drawing on the results of that study, Magistrate Judge Andrew Peck, in an opinion with far-reaching consequences, gave court approval for the use of A/IS to conduct legal discovery.[65]

The story of the TREC Legal Track's role in facilitating the adoption of A/IS for legal fact-finding contains a few lessons:

- **Metrics:** By focusing on recall and precision, the TREC studies quantified the effectiveness of the systems evaluated in a way that legal practitioners could readily understand.
- **Benchmarks:** The TREC studies filled an important gap: independent, scientifically sound evaluations of the effectiveness of A/IS applied to the real-world challenge of legal e-discovery.
- **Collaboration:** The founders of the TREC studies and the most successful participants came from both scientific and legal backgrounds, demonstrating the importance of multidisciplinary collaboration.

The TREC studies are a shining example of how the truth-seeking protocols of science can be used to advance the truth-seeking protocols of the law. They can serve as a conceptual basis for future benchmarking efforts, as well as the development of standards and certification programs to support informed trust when it comes to effectiveness of A/IS deployed in legal systems.[66]

Law

Recommendations

1. Governments should fund and support the establishment of ongoing benchmarking exercises designed to provide valid, publicly accessible measurements of the effectiveness of A/IS deployed, or potentially deployed, in the legal system. That support could take a number of forms, ranging from direct sponsorship and oversight—for example, by nonregulatory measurement laboratories such as the US NIST—to indirect support by the recognition of the results of a credible third-party benchmarking exercise for the purposes of meeting procurement and contracting requirements. All government efforts in this regard should be transparent and open to public scrutiny.
2. Governments should facilitate the creation of data sets that can be used for purposes of evaluating the effectiveness of A/IS as applied in the legal system. In assisting in the creation of such data sets, governments and administrative agencies will have to take into consideration potentially competing societal values, such as the protection of personal data, and arrive at solutions that maintain those values while enabling the creation of usable, real-world data sets. All government efforts in this regard should be transparent and open to public scrutiny.
3. Creators of A/IS to be applied to legal matters should pursue valid measures of the effectiveness of their systems, whether through participation in benchmarking exercises or through conducting single-system validation exercises. Creators should describe the procedures and results of the testing in clear language that is understandable to both experts and nonexperts, and should do so without disclosing intellectual property. Further, the descriptions should be open to examination by all stakeholders, including, when appropriate, the general public.
4. Researchers engaged in the study and development of A/IS for use in the legal system should seek to define meaningful metrics that gauge the effectiveness of the systems they study. In selecting and defining metrics, researchers should seek input from all stakeholders in the outcome of the given application of A/IS in the legal system. The metrics should be readily understandable by experts and nonexperts alike.
5. Governments and industry associations should undertake educational efforts to inform both those engaged in the operation of A/IS deployed in the legal system and those affected by the results of their operation of the salient measures of effectiveness and what they can indicate about the capabilities and limitations of the A/IS in question.
6. Creators of A/IS for use in the legal system should ensure that the effectiveness metrics defined by the research community are readily obtainable and accessible to all stakeholders, including, when appropriate, the general public. Creators should provide guidance on how to interpret and respond to the metrics generated by the system.
7. Operators of A/IS applied to a legal task should follow the guidance on the measurement of effectiveness provided for

the A/IS being used. This includes guidance about which metrics to obtain, how and when to obtain them, how to respond to given results, when it may be appropriate to follow alternative methods of gauging effectiveness, and so on.

8. In interpreting and responding to measurements of the effectiveness of A/IS applied to legal problems or questions, allowance should be made by those interpreting the results for variation in the specific objectives and circumstances of a given deployment of A/IS. Quantitative results should be supplemented by qualitative evaluation of the practical significance of a given outcome and whether it indicates a need for remediation. This evaluation should be done by an individual with the technical expertise and pragmatic experience needed to make a sound judgment.
9. Industry associations or other organizations should collaborate on developing standards for measuring and reporting on the effectiveness of A/IS. These standards should be developed with input from both the scientific and legal communities.
10. Recommendation 1 under Issue 2, with respect to effectiveness.
11. Recommendation 2 under Issue 2, with respect to effectiveness.

Further Resources

- *Da Silva Moore v. Publicis Groupe*, 2012 WL 607412 (S.D.N.Y. Feb. 24, 2012).
- C. Garvie, A. M. Bedoya, and J. Frankle, "The Perpetual Line-Up: Unregulated Police Face Recognition in America," Georgetown Law, Center on Privacy & Technology, Oct. 2016.
- M. R. Grossman and G. V. Cormack, "Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review," Richmond Journal of Law and Technology, vol. 17, no. 3, 2011.
- B. Hedin, D. Brassil, and A. Jones, "On the Place of Measurement in E-Discovery," in Perspectives on Predictive Coding and Other Advanced Search Methods for the Legal Practitioner, J. R. Baron, R. C. Losey, and M. D. Berman, Eds. Chicago: American Bar Association, 2016.
- J. A. Kroll, "The fallacy of inscrutability," Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences, vol. 376, no. 2133, Oct. 2018.
- D. W. Oard, J. R. Baron, B. Hedin, D. Lewis, and S. Tomlinson, "Evaluation of Information Retrieval for E-Discovery," Artificial Intelligence and Law, vol. 18, no. 4, pp. 347-386, Aug. 2010.
- The Sedona Conference, "The Sedona Conference Commentary on Achieving Quality in the E-Discovery Process," The Sedona Conference Journal, vol. 15, pp. 265-304, 2014.
- M. T. Stevenson, "Assessing Risk Assessment in Action," Minnesota Law Review, vol. 103, June 2018.

- "Global Governance of AI Roundtable: Summary Report 2018," World Government Summit, 2018.
- High-Level Expert Group on Artificial Intelligence, "DRAFT Ethics Guidelines for Trustworthy AI: Working Document for Stakeholders' Consultation," The European Commission. Brussels, Belgium: Dec. 18, 2018.

Issue 4: Competence

How can specification of the knowledge and skills required of the human operator(s) of A/IS foster informed in the suitability of A/IS for adoption in legal systems?

Background

An essential component of informed trust in a technological system, especially one that may affect us in profound ways, is confidence in the competence of the operator(s) of the technology. We trust surgeons or pilots with our lives because we have confidence that they have the knowledge, skills, and experience to apply the tools and methods needed to carry out their tasks effectively. We have that confidence because we know that these operators have met rigorous professional and scientific accreditation standards before being allowed to step into the operating room or cockpit. This informed trust in operator competence is what gives us confidence that surgery or air travel will result in the desired outcome. No such standards of operator competence currently exist with respect to A/IS applied in legal systems, where the life, liberty, and rights of citizens can be at stake. That absence of standards hinders the trustworthy adoption of A/IS in the legal domain.

The human operator is an integral component of A/IS

Almost all current applications of A/IS in legal systems, like those in most other fields, require human mediation and likely will continue to do so for the near future. This human mediation, post design and post development, will take a number of forms, including decisions about (a) whether or not to use A/IS for a given purpose,[67] (b) the data used to train the systems, (c) settings for system parameters to be used in generating results, (d) methods of validating results, (e) interpretation and application of the results, and so on. Because these systems' outcomes are a function of all their components, including the human operator(s), their effectiveness, and by extension trustworthiness, will depend on their human operator(s).

Despite this, there are few standards that specify how humans should mediate applications of A/IS in legal systems, or what knowledge qualifies a person to apply A/IS and interpret their results.[68] This reality is especially troubling for the instances in which the life, rights, or liberty of humans are at stake. Today, while professional codes of ethics for lawyers are beginning to include among their

Law

requirements an awareness and understanding of technologies with legal application,[69] the operators of A/IS in legal systems are essentially deemed to be capable of determining their own competence: lawyers or IT professionals operating in civil discovery, correctional officers using risk assessment algorithms, and law enforcement agencies engaging in predictive policing or using automated surveillance technologies. All are mostly able to use A/IS without demonstrating that they understand the operation of the system they are using or that they have any particular set of consensus competencies.[70]

The lack of competency requirements or standards undermines the establishment of informed trust in the use of A/IS in legal systems. If courts, legal practitioners, law enforcement agencies, and the general public are to rely on the results of A/IS when applied to tasks traditionally carried out by legal professionals, they must have grounds for believing that those operating A/IS will possess the requisite knowledge and skill to understand the conditions and methods for operating the systems effectively, including evaluating the data on which the A/IS trained, the data to which they are applied, the results they produce, and the methods and results of measuring the effectiveness the systems. Applied incompetently, A/IS could produce the opposite intended effect. Instead of improving a legal system—and bringing about the gains in well-being that follow from such improvements—they may undermine both the fairness and effectiveness of a legal system and trust in its fairness and effectiveness, creating conditions for social disorder and the deterioration of human well-being that would follow from that disorder. By way of illustration:

- A city council might misallocate funds for policing across city neighborhoods because it relies on the output of an algorithm that directs attention to neighborhoods based on arrest rates rather than actual crime rates.[71]
- In civil justice, A/IS applied in a search of documents to uncover relevant facts may fail to do so because an operator without sufficient competence in statistics may materially overestimate the accuracy of the system, thus ceasing vital fact-finding activities.[72]
- In the money bail system, reliance on A/IS to reduce bias may instead perpetuate it. For example, if a judge does not understand whether an algorithm makes sufficient contextual distinctions between gradations of offenses,[73] that judge would not able to probe the output of the A/IS and make a well-informed use of it.
- In the criminal justice system, an operator using A/IS in sentencing decision-support may fail to identify bias, or to assess the risk of bias, in the results generated by the A/IS,[74] unfairly depriving a citizen of his or her liberty or prematurely granting an offender's release, increasing the risk of recidivism.

More generally, without the confidence that A/IS operators will apply the technology as intended and supervise it appropriately, the general public will harbor fear, uncertainty, and doubt about the use of A/IS in legal systems and potentially about the legal systems themselves.

Law

Fostering informed trust in the competence of human operators

If negative outcomes such as those just described are to be avoided, **it will be necessary to include among norms for the adoption of A/IS in a legal system a provision for building informed trust in the operators of A/IS**. Building trust will require articulating standards and best practices for two groups of agents involved in the deployment of A/IS: creators and operators.

On the one hand, those engaged in the design, development, and marketing of A/IS must commit to specifying the knowledge, skills, and conditions required for the safe, ethical, and effective deployment and operation of the systems.[75] On the other hand, those engaged in actually operating the systems, including both legal professionals and experts acting in the service of legal professionals, must commit to adhering to these requirements in a manner consistent with other operative legal, ethical, and professional requirements. The precise nature of the competency requirements will vary with the nature and purpose of the A/IS and what is at stake in their effective operation. The requirements for the operation of A/IS designed to assist in the creation of contracts, for example, might be less stringent than those for the operation of A/IS designed to assess flight risk, which could affect the liberty of individual citizens.

A corollary of these provisions is that education and training in the requisite skills should be available and accessible to those who would operate A/IS, whether that training is provided through professional schools, such as law school; through institutions providing ongoing professional training, such as, for federal judges in the United States, the Federal Judicial Center; through professional and industry associations, such as the American Bar Association; or through resources accessible by the general public.[76] Making sure such training is available and accessible will be essential to ensuring that the resources needed for the competent operation of A/IS are widely and equitably distributed.[77]

It will take a combined effort of both creators and operators to ensure both that A/IS designed for use in legal systems are properly applied and that those with a stake in the effective functioning of legal systems—including legal professionals, of course, but also decision subjects, victims of crime, communities, and the general public—will have informed trust, or, for that matter, informed distrust (if that is what a competence assessment finds) in the competence of the operators of A/IS as applied to legal problems and questions.[78]

Illustration—Competence

Included among the offerings of Amazon Web Services is an image and video analysis service known as Amazon Rekognition.[79] The service is designed to enable the recognition of text, objects, people, and actions in images and videos. The technology also enables the search and comparison of faces, a feature with potential law enforcement and national security applications, such as comparing faces identified in video taken by a security camera with those in a database of jail booking photos. Attracted by

Law

the latter feature, police departments in Oregon and Florida have undertaken pilots of Rekognition as a tool in their law enforcement efforts.[80]

In 2018, the American Civil Liberties Union (ACLU), a frequent critic of the use of facial recognition technologies by law enforcement agencies,[81] conducted a test of Rekognition. The test consisted of first constructing a database of 25,000 booking photos ("mugshots") then comparing publicly available photos of all then-current members of the US Congress against the images in the database. The test found that Rekognition incorrectly matched the faces of 28 members of Congress with faces of individuals who had been arrested for a crime.[82] The ACLU argues that the high number of false positives generated by the technology shows that police use of facial recognition technologies generally (and of Rekognition in particular) poses a risk to the privacy and liberty of law-abiding citizens. The ACLU has used the results of its test of Rekognition to support its proposal that Congress enact a moratorium on the use of facial recognition technologies by law enforcement agencies until stronger safeguards against their misuse, and potential abuse, can be put in place.[83]

In response to the ACLU report, Amazon noted that the ACLU researchers, in conducting their study, had applied the technology utilizing a similarity threshold (a gauge of the likelihood of a true match) of 80%, a threshold that casts a fairly wide net for potential matches (and hence generates a high number of false positives). For applications in which there are greater costs associated with false positives (e.g., policing), Amazon recommends utilizing a similarity threshold value of 99% or above to reduce accidental misidentification.[84] Amazon also noted that, in all law enforcement use cases, it would be expected that the results of the technology would be reviewed by a human before any actual police action would be undertaken.

The story of the ACLU's testing of Rekognition and Amazon's response to the test highlights the importance of specifying and adhering to guidelines for competent use.[85] Had a law enforcement agency used the technology in the way it was used in the ACLU test, it would, in most legitimate use cases, be guilty of incompetent use. At the same time, Amazon is not free of blame insofar as it did not specify prominently and clearly the competency guidelines for effective use of the technology in support of law enforcement efforts, as well as the risks that might be incurred if those guidelines are not followed. Competent use[86] follows both from the A/IS creator's specification of well-grounded[87] competency guidelines and from the A/IS operator's adherence to those guidelines.[88]

Recommendations

1. Creators of A/IS for application in legal systems should provide clear and accessible guidance for the knowledge, skills, and experience required of the human operators of the A/IS if the systems are to achieve expected levels of effectiveness. Included in that guidance should be a delineation of the risks involved if those requirements are not met. Such guidance should be

Law

documented in a form that is accessible and understandable by both experts and the general public.

2. Creators and developers of A/IS for application in legal systems should create written policies that govern how the A/IS should be operated. In creating these policies, creators and developers should draw on input from the legal professionals who will be using the A/IS they are creating. The policies should include:
   - the specification of the real-world applications for the A/IS;
   - the preconditions for their effective use;
   - the training and skills that are required for operators of the systems;
   - the procedures for gauging the effectiveness of the A/IS;
   - the considerations to take into account in interpreting the results of the A/IS;
   - the outcomes that can be expected by both operators and other affected parties when the A/IS are operated properly; and
   - the specific risks that follow from improper use.

   The policies should also specify circumstances in which it might be necessary for the operator to override the A/IS. All such policies should be publicly accessible.

3. Creators and developers of A/IS to be applied in legal systems should integrate safeguards against the incompetent operation of their systems. Safeguards could include issuing notifications and warnings to operators in certain conditions, requiring, as appropriate, acknowledgment of receipt; limiting access to A/IS functionality based on the operator's level of expertise; enabling system shut-down in potentially high-risk conditions; and more. These safeguards should be flexible and governed by context-sensitive policies set by competent personnel of the entity (e.g., the judiciary), utilizing the A/IS to address a legal problem.

4. Governments should provide that any individual whose legal outcome is affected by the application of A/IS should be notified of the role played by A/IS in that outcome. Further, the affected party should have recourse to appeal to the judgment of a competent human being.

5. Professionals engaged in the creation, practice, interpretation, and enforcement of the law, such as lawyers, judges, and law enforcement officers, should recognize the specialized scientific and professional expertise required for the ethical and effective application of A/IS to their professional duties. The professional associations to which such legal practitioners belong, such as the American Bar Association, should, through both educational programs and professional codes of ethics, seek to ensure that their members are well informed about the scientific and technical competency requirements for the effective and trustworthy application of A/IS to the law.[89]

6. The operators of A/IS applied in legal systems—whether the operator is a specialist in A/IS or a legal professional—should

understand the competencies required for the effective performance of their roles and should either acquire those competencies or identify individuals with those competencies who can support them in the performance of their roles. The operator does not need to be an expert in all the pertinent domains but should have access to individuals with the requisite expertise.

7. Recommendation 1 under Issue 2, with respect to competence.
8. Recommendation 2 under Issue 2, with respect to competence.

Further Resources

- C. Garvie, A. M. Bedoya, and J. Frankle, "The Perpetual Line-Up: Unregulated Police Face Recognition in America," Georgetown Law, Center on Privacy & Technology, Oct. 2016.
- International Organization for Standardization, *ISO/IEC 27050-3: Information technology—Security techniques—Electronic discovery—Part 3: Code of practice for electronic discovery*, Geneva, 2017.
- J. A. Kroll, "The fallacy of inscrutability," Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences, vol. 376, no. 2133, Oct. 2018.
- A. G. Ferguson, "Policing Predictive Policing," Washington University Law Review, vol. 94, no. 5 2017.
- "Global Governance of AI Roundtable: Summary Report 2018," World Government Summit, 2018.

Issue 5: Accountability

How can the ability to apportion responsibility for the outcome of the application of A/IS foster informed trust in the suitability of A/IS for adoption in legal systems?

Background

Apportioning responsibility. An essential component of informed trust in a technological system is confidence that it is possible, if the need arises, to apportion responsibility among the human agents engaged along the path of its creation and application: from design through to development, procurement, deployment,[90] operation, and, finally, validation of effectiveness. Unless there are mechanisms to hold the agents engaged in these steps accountable, it will be difficult or impossible to assess responsibility for the outcome of the system under any framework, whether a formal legal framework or a less formal normative framework. A model of A/IS creation and use that does not have such mechanisms will also lack important forms of deterrence against poorly thought-out design, casual adoption, and inappropriate use of A/IS.

Simply put, a system that produces outcomes for which no one is responsible cannot be trusted. Those engaged in creating, procuring, deploying, and operating such a system will lack the discipline engendered by the clear

assignment of responsibility. Meanwhile, those affected by the results of the system's operation will find their questions around a given result inadequately answered, and errors generated by the system will go uncorrected. In the case of A/IS applied in a legal system, where an individual's basic human rights may be at issue, these questions and errors are of fundamental importance. In such circumstances, the only options are either blind trust or blind distrust. Neither of those options is satisfactory, especially in the case of a technological system applied in a domain as fundamental to the social order as the law.

Challenges to accountability

In the case of A/IS, whether applied in a legal system or another domain, maintaining accountability can be a particularly steep challenge. This challenge to accountability is because of both the perceived "black box" nature of A/IS and the diffusion of responsibility it brings.

The perception of A/IS as a black box stems from the opacity that is an inevitable characteristic of a system that is a complex nexus of algorithms, computer code, and input data. As observed by Joshua New and Daniel Castro of the Information Technology and Innovation Foundation:

The most common criticism of algorithmic decision-making is that it is a "black box" of extraordinarily complex underlying decision models involving millions of data points and thousands of lines of code. Moreover, the model can change over time, particularly when using machine learning algorithms that adjust the model as the algorithm encounters new data.[91]

This opacity of the systems makes it challenging to trace cause to effect,[92] which, in turn, makes it difficult or even impossible, to draw lines of responsibility.

The diffuseness challenge stems from the fact that even the most seemingly straightforward A/IS can be complex, with a wide range of agents—systems designers, engineers, data analysts, quality control specialists, operators, and others—involved in design, development, and deployment. Moreover, some of these agents may not even have been engaged in the development of the A/IS in question; they may have, for example, developed open-source components that were intended for an entirely different purpose but that were subsequently incorporated into the A/IS. This diffuseness of responsibility poses a challenge to the maintenance of accountability.[93] As Matthew Scherer, a frequent writer and speaker on topics at the intersection of law and A/IS, observes:

The sheer number of individuals and firms that may participate in the design, modification, and incorporation of an AI system's components will make it difficult to identify the most responsible party or parties. Some components may have been designed years before the AI project had even been conceived, and the components' designers may never have envisioned, much less intended, that their designs would be incorporated into any AI system, still less the specific AI system that caused harm. In such circumstances, it may seem unfair to assign

blame to the designer of a component whose work was far-removed in both time and geographic location from the completion and operation of the AI system.[94]

Examples include the following:

- When a judge's ruling includes a long prison sentence, based in part on a flawed A/IS-enabled process that erroneously deemed a particular person to be at high risk of recidivism, who is responsible for the error? Is it the A/IS designer, the person who chose the data or weighed the inputs, the prosecution team who developed and delivered the risk profile to the court, or the judge who did not have the competence to ask the appropriate questions that would have enabled a clearer understanding of the limitations of the system? Or is responsibility somehow distributed among these various agents?[95]
- When a lawyer engaged in civil or criminal discovery believes, erroneously, that all the relevant information was found when using A/IS in a data-intensive matter, who is responsible for the failure to gather important facts? The A/IS designer who typically would have had no ability to foretell the specific circumstances of a given matter, the legal or IT professional who operated the A/IS or erroneously measured its effectiveness, or the lawyer who made a representation to his or her client, to the court, or to investigatory agencies?
- When a law enforcement officer, relying on A/IS, erroneously identifies an individual as being more likely to commit a crime than another, who is responsible for the resulting encroachment on the civil rights of the person erroneously targeted? Is it the A/IS designer, the individual who selected the data on which to train the algorithm, the individual who chose how the effectiveness of the A/IS would be measured,[96] the experts who provided training to the officer, or the officer himself or herself?

As a result of the challenges presented by the opacity and diffuseness of responsibility in A/IS, the present-day answer to the question, "Who is accountable?" is, in far too many instances, "It's hard to say." This is a response that, in practice, means "no one" or, equally unhelpful, "everyone". Such failure to maintain accountability will undermine efforts to bring A/IS (and all their potential benefits) into legal systems based on informed trust.

Maintaining accountability and trust in A/IS

Although maintaining accountability in complex systems can be a challenge, it is one that must be met in order to engender informed trust in the use of A/IS in the legal domain. "Blaming the algorithm" is not a substitute for taking on the challenge of maintaining transparent lines of responsibility and establishing norms of accountability.[97] This is true even if we allow that, given the complexity of the systems in question, some number of "systems accidents" is inevitable.[98] Informed trust in a system does not require a belief that zero errors will occur; however, it does require a belief that there are mechanisms in place for addressing errors when

Law

they do occur. Accountability is an essential component of those mechanisms.

In meeting the challenge, it should be recognized that there are existing norms and controls that have a role to play in ensuring that accountability is maintained. For example, contractual arrangements between the A/IS provider and a party acquiring and applying a system may help to specify who is (and is not) to be held liable in the event the system produces undesirable results. Professional codes of ethics may also go some way toward specifying the extent to which lawyers, for example, are responsible for the results generated by the technologies they use, whether they operate them directly or retain someone else to do so. Judicial systems may have procedures for assessing responsibility when a citizen's rights are improperly infringed. As illustrated by the cases described above, however, existing norms and controls, while helpful, are insufficient in themselves to meet the specific challenge represented by the opacity and diffuseness of A/IS. To meet the challenge further steps must be taken.[99]

The first step is ensuring that all those engaged in the creation, procurement, deployment, operation, and testing of A/IS recognize that, if accountability is not maintained, these systems will not be trusted. In the interest of maintaining accountability, these stakeholders should take steps to clarify lines of responsibility throughout this continuum, and make those lines of responsibility, when appropriate, accessible to meaningful inquiry and audit.

The goal of clarifying lines of responsibility in the operation of A/IS is to implement a governing model that specifies who is responsible for what, and who has recourse to which corrective actions, i.e., a trustworthy model that ensures that it will admit actionable answers should questions of accountability arise. Arriving at an effective model will require the participation of those engaged in the creation and operation of A/IS, those affected by the results of their use, and those with the expertise to understand how such a model would be used in a given legal system. For example:

- Individuals responsible for the design of A/IS will have to maintain a transparent record of the sources of the various components of their systems, including identification of which components were developed in-house and which were acquired from outside sources, whether open source or acquired from another firm.
- Individuals responsible for the design of A/IS will have to specify the roles, responsibilities, and potential subsequent liabilities of those who will be engaged in the operation of the systems they create.
- Individuals responsible for the operation of a system will have to understand their roles, responsibilities, potential liabilities, and will have to maintain documentation of their adherence to requirements.
- Individuals affected by the results of the operation of A/IS, e.g., a defendant in a criminal proceeding, will have to be given access to information about the roles and responsibilities of those involved in relevant

Law

aspects of the creation, operation, and validation of the effectiveness of the A/IS affecting them.[100]

- Individuals with legal and political training (e.g., jurists, regulators, as well as legal and political scholars) will have to ensure that any model that is created will provide information that is in fact actionable within the operative legal system.

A governing model of accountability that reflects the interests of all these stakeholders will be more effective both at deterring irresponsible design or use of A/IS before it happens and at apportioning responsibility for an undesirable outcome when it does happen.[101]

Pulling together the input from the various stakeholders will likely not take place without some amount of institutional initiative. Organizations that employ A/IS for accomplishing legal tasks—private firms, regulatory agencies, law enforcement agencies, judicial institutions—should therefore develop and implement policies that will advance the goal of clarifying lines of responsibility. Such policies could take the form of, for example, designating an official specifically charged with oversight of the organization's procurement, deployment, and evaluation of A/IS as well as the organization's efforts to educate people both inside and outside the organization on its use of A/IS. Such policies might also include the establishment of a review board to assess the organization's use of A/IS and to ensure that lines of responsibility for the outcomes of its use are maintained. In the case of agencies, such as police departments, whose use of A/IS could impact the general public, such review boards would, in the interest of legitimacy, have to include participation from various citizens' groups, such as those representing defendants in the criminal system as well as those representing victims of crime.[102]

The goal of opening lines of responsibility to meaningful inquiry is to ensure that an investigation into the use of A/IS will be able to isolate responsibility for errors (or potential errors) generated by the systems and their operation.[103] This means that all those engaged in the design, development, procurement, deployment, operation, and validation of the effectiveness of A/IS, as well as the organizations that employ them, must in good faith be willing to participate in an audit, whether the audit is a formal legal investigation or a less formal inquiry. They must also be willing to create and preserve documentation of key procedures, decisions, certifications,[104] and tests made in the course of developing and deploying the A/IS.[105]

The combination of a governing model of accountability and an openness to meaningful audit will allow the maintenance of accountability, even in complex deployments of A/IS in the service of a legal system.

Additional note 1. The principle of accountability is closely linked with each of the other principles intended to foster informed trust in A/IS: effectiveness, competence, and transparency. With respect to effectiveness, evidence of attaining key metrics and benchmarks to confirm that A/IS are functioning as intended may put questions of where, among creators,

owners, and operators, responsibility for the outcome of a system lies on a sound empirical footing. With respect to competence, operator credentialing and specified system handoffs enable a clear chain of responsibility in the deployment of A/IS.[106] With respect to transparency, providing a view into the general design and methods of A/IS, or even a specific explanation for a given outcome, can help to advance accountability.

Additional note 2. Closely related to accountability is the trust that follows from knowing that a human expert is guiding the A/IS and is capable of overriding them, if necessary. Subjecting humans to automated decisions not only raises legal and ethical concerns, both from a data protection[107] and fundamental rights perspective,[108] but also will likely be viewed with distrust if the human component, which can introduce circumstantial flexibility in the interest of realizing an ethically superior outcome, is missing. In addition to ensuring technical safety and reliability of A/IS used in the course of decision-making processes, the legal system should also, where appropriate, provide for the possibility of an appeal for review by a human judge. Careful attention must be paid to the design of corresponding appeal procedures.[109]

Illustration—Accountability

Over the last two decades, criminal justice agencies have increasingly embraced predictive tools to assist in the determination for bail, sentencing, and parole. A mix of companies, government agencies, nonprofits, and universities have built and promoted tools that provide a likelihood that someone may fail to appear or may commit a new crime or a new violent act. While math has played a role in these determinations since at least the 1920s,[110] a new interest in accountability and transparency has brought novel legal challenges to these tools.

In 2013, Eric Loomis was arrested for a drive-by shooting in La Crosse, Wisconsin. No one was hit, but Loomis faced prison time. Loomis denied involvement in the shooting, but waived his right to trial and entered a guilty plea to two of the less severe offenses with which he was charged: attempting to flee a traffic officer and operating a motor vehicle without the owner's consent. The judge sentenced him to six years in prison, saying he was "high risk". The judge based this conclusion, in part, on the risk assessment score given by Compas, a secret and privately held algorithmic tool used routinely by the Wisconsin Department of Corrections.

On appeal, Loomis made three major arguments, two focused on accountability.[111] First, the tool's proprietary nature—the underlying code was not made available to the defense—made it impossible to test its scientific validity. Second, the tool inappropriately considered gender in making its determination.

A unanimous Wisconsin Supreme Court ruled against Loomis on both arguments.

The court reasoned that knowing the inputs and output of the tool, and having access to validating studies of the tool's accuracy, were sufficient to prevent infringement of Loomis' due process.[112] Regarding the use of gender—a protected class in the United States—the court said he did not show that there was a reliance on gender in making the output or sentencing decision.

Law

Without the ability to interrogate the tool and know how gender is used, the court created a paradox with its opinion.

The *Loomis* decision represents the challenges that judges have balancing accountability of "black boxed" A/IS and trade secret protections.[113] Other decisions have sided against accountability of other risk assessments,[114] probabilistic DNA analysis tools,[115] and government remote hacking investigation software.[116] Siding with accountability, a federal judge found that the underlying code of a probability software used in DNA comparisons was admissible and relevant to a pretrial hearing where the admissibility of expert testimony is challenged.[117]

These issues will continue to be litigated as A/IS tools continue to proliferate in judicial systems. To that end, as the *Loomis* court notes, "The justice system must keep up with the research and continuously assess the use of these tools."

Recommendations

1. Creators of A/IS to be applied in a legal system should articulate and document well-defined lines of responsibility, among all those who would be engaged in the development and operation of the A/IS, for the outcome of the A/IS.
2. Those engaged in the adoption and operation of A/IS to be applied in a legal system should understand their specific responsibilities for the outcome of the A/IS as well as their potential liability should the A/IS produce an outcome other than that intended. In the case of A/IS, many questions of legal liability remain unsettled. Adopters and operators of A/IS should nevertheless understand to what extent they could, *potentially*, be held liable for an undesirable outcome.
3. When negotiating contracts for the provision of A/IS products and services for use in the legal system, providers and buyers of A/IS should include contractual terms specifying clear lines of responsibility for the outcomes of the systems being acquired.
4. Creators and operators of A/IS applied in a legal system, and the organizations that employ them, should be amenable to internal oversight mechanisms and inquiries (or audits) that have the objective of allocating responsibility for the outcomes generated by the A/IS. In the case of A/IS adopted and deployed by organizations that have direct public interaction (e.g., a law enforcement agency), oversight and inquiry could also be conducted by external review boards. Being prepared for such inquiries means maintaining clear documentation of all salient procedures followed, decisions made, and tests conducted in the course of developing and applying the A/IS.
5. Organizations engaged in the development and operation of A/IS for legal tasks should consider mechanisms that will create individual and collective incentives for ensuring both that the outcomes of the A/IS adhere to ethical standards and that accountability for those outcomes is maintained, e.g., mechanisms to ensure that speed and efficiency are not rewarded at the expense of a loss of accountability.

6. Those conducting inquiries to determine responsibility for the outcomes of A/IS applied in a legal system should take into consideration all human agents involved in the design, development, procurement, deployment, operation, and validation of effectiveness of the A/IS and should assign responsibility accordingly.
7. Recommendation 1 under Issue 2, with respect to accountability.
8. Recommendation 2 under Issue 2, with respect to accountability.

Further Resources

- N. Diakopoulos, S. Friedler, M. Arenas, S. Barocas, M. Hay, B. Howe, H. V. Jagadish, K. Unsworth, A. Sahuguet, S. Venkatasubramanian, C. Wilson, C. Yu, and B. Zevenbergen, "Principles for Accountable Algorithms and a Social Impact Statement for Algorithms," FAT/ML.
- F. Doshi-Velez, M. Kortz, R. Budish, C. Bavitz, S. J. Gershman, D. O'Brien, S. Shieber, J. Waldo, D. Weinberger, and A. Wood, "Accountability of AI Under the Law: The Role of Explanation," Berkman Center Research Publication Forthcoming; Harvard Public Law Working Paper, no. 18-07, Nov. 3, 2017.
- European Commission for the Efficiency of Justice. *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*. Strasbourg, 2018.
- J. A. Kroll, J. Huey, S. Barocas, E. W. Felten, J. R. Reidenberg, D. G. Robinson, and H. Yu, "Accountable Algorithms," University of Pennsylvania Law Review, vol. 165, pp. 633-705. Feb. 2017.
- J. New and D. Castro, "How Policymakers Can Foster Algorithmic Accountability," Information Technology and Innovation Foundation, May 21, 2018.
- M. U. Scherer, "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies," Harvard Journal of Law & Technology, vol. 29. no. 2, pp. 369-373, 2016.
- J. Tashea, "Calculating Crime: Attorneys are Challenging the Use of Algorithms to Help Determine Bail, Sentencing and Parole," ABA Journal, March 2017.

Issue 6: Transparency

How can sharing information that explains how A/IS reached given decisions or outcomes foster informed trust in the suitability of A/IS for adoption in legal systems?

Background

Access to meaningful information.
An essential component of informed trust in a technological system is confidence that the information required for a human to understand why the system behaves a certain way in a specific circumstance (or would behave in

a hypothetical circumstance) will be accessible. Without transparency, there is no basis for trusting that a given decision or outcome of the system can be explained, replicated, or, if necessary, corrected.[118] Without transparency, there is no basis for informed trust that the system can be operated in a way that achieves its ends reliably and consistently or that the system will not be used in a way that impinges on human rights. In the case of A/IS applied in a legal system, such a lack of trust could undermine the credibility of the legal system itself.

Transparency and trust

Transparency, by prioritizing access to information about the operation and effectiveness of A/IS, serves the purpose of fostering informed trust in the systems. More specifically, transparency fosters trust that:

- the operation of A/IS and the results they produce are explainable;
- the operation and results of A/IS are fair;[119]
- the operation and results of A/IS are unbiased;
- the A/IS meet normative standards for operation and results;
- the A/IS are effective;
- the results of A/IS are replicable;[120] and
- those engaged in the design, development, procurement, deployment, operation, and validation of the effectiveness of A/IS can be held accountable, where appropriate, for negative outcomes, and that corrective or punitive action can be taken when warranted.

For A/IS used in a legal system to achieve their intended purposes, all those with a stake in the effective functioning of the legal system must have a well-grounded trust that the A/IS can meet these requirements. This trust can be fostered by transparency.

The elements of transparency

Transparency of A/IS in legal matters requires disclosing information about the design and operation of the A/IS to various stakeholders. In implementing the principle, however, we must, in the interest of both feasibility and effectiveness, be more precise both about the categories of stakeholders to whom the information will be disclosed, and about the categories of information that will be disclosed to those stakeholders.

Relevant stakeholders in a legal system include those who:

- operate A/IS for the purpose of carrying out tasks in civil justice, criminal justice, and law enforcement, such as a law enforcement officer who uses facial recognition tools to identify potential suspects;
- rely on the results of A/IS to make important decisions, such as a judge who draws on the results of an algorithmic assessment of recidivism risk in deciding on a sentence;
- are directly affected by the use of A/IS—a "decision subject", such as a defendant in a criminal proceeding whose bail terms are influenced by an algorithmic assessment of flight risk;

Law

- are indirectly affected by the results of A/IS, such as the members of a community that receives more or less police attention because of the results of predictive policing technology; and
- have an interest in the effective functioning of the legal system, such as judges, lawyers, and the general public.

Different types of relevant information can be grouped into high-level categories. As illustrated below, a taxonomy of such high-level categories may, for example, distinguish between:

- nontechnical procedural information regarding the employment and development of a given application of A/IS;
- information regarding data involved in the development, training, and operation of the system;
- information concerning a system's effectiveness/performance;
- information about the formal models that the system relies on; and
- information that serves to explain a system's general logic or specific outputs.

These more granular distinctions matter because different sorts of inquiries will require different sorts of information, and it is important to match the information provided to the actual needs of the inquiry. For example, an inquiry into a predictive policing system that misdirected police resources may not be much advanced by information about the formal models on which the system relied, but it may well be advanced by an explanation for the specific outcome.

On the other hand, an inquiry, undertaken by a designer or operator, into ways to improve system performance may benefit from access to information about the formal models on which the system relies.[121]

These distinctions also matter because there may be circumstances in which it would be desirable to limit access to a given type of information to certain stakeholders. For example, there may be circumstances in which one would want to identify an agent to serve as a public interest steward. For auditing purposes, this individual would have access to certain types of sensitive information unavailable to others. Such restrictions on information access are necessary if the transparency principle is not to impinge on other societal values and goals, such as security, privacy, and appropriate protection of intellectual property.[122]

The salience of the question, "*Who is given access to what information?*" is illustrated by Sentiment Meter, a technology developed by Elucd, a GovTech company that provides cities with near real-time understanding of how citizens feel about their government, in conjunction with the New York Police Department, to assist the NYPD in gauging citizens' views regarding police activity in their communities.[123] One of the stated goals of the program is to build public trust in the police department. In the interest of trust, should "the public" have access to all potentially relevant information, including how the system was designed and developed, what the input data are, who operates the system and what their qualifications are, how the system's effectiveness was tested, and why the public was not brought

into the process of construction? If the answer is that the general public should not have access to all this information, then who should? How do we define "the public?" Is it the whole community represented in its elected officials? Or should certain communities have greater access, for example, those most affected by controversial police practices such as stop, question, and frisk? Such questions must be answered if the program is to achieve its stated goals.

Transparency in practice

As just noted, although transparency can foster informed trust in A/IS applied in a legal system, **its practical implementation requires careful thought.** Requiring public access to all information pertaining to the operation and results of A/IS is neither necessary nor feasible. What is required is a careful consideration of who needs access to what information for the specific purpose of building informed trust. The following table is an example of a tool that might be used to match type of information to type of information consumer for the purpose of fostering trust.[124]

Law

Types of information that should be considered in determining transparency demands in relation to a given A/IS		Stakeholders whose interest in access to different types of information should be considered in determining the transparency demands in relation to a given application of A/IS			
High-level category	Specific type of information (examples) Disclosure of...	Operators	Decision-subjects	Public interest steward	General public
Procedural aspects regarding A/IS employment and development	the fact that a given context involves the employment of A/IS	N/A	?	?	?
	how the employment of the system was authorized	?	?	?	?
	who developed the system	?	?	?	?
	...				
Data involved in A/IS development and operation	the origins of training data and data involved in the operation of the system	?	?	?	?
	the kinds of quality checks that data was subject to and their results	?	?	?	?
	how data labels are defined and to what extent data involves proxy variables	?	?	?	?
	relevant data sets themselves	?	?	?	?
	...				
Effectiveness/ performance	the kinds of effectiveness/performance measurement that have occurred	?	?	?	?
	measurement results	?	?	?	?
	any independent auditing or certification	?	?	?	?
	...				
Model specification	the input variables involved	?	?	?	?
	the variable(s) that the model optimizes for	?	?	?	?
	tthe complete model (complete formal representation, source code, etc.)	?	?	?	?
	...				
Explanation	information concerning the system's general logic or functioning	?	?	?	?
	information concerning the determinants of a particular output [125]	?	?	?	?
	...				

Law

When it comes to deciding whether a specific type of information should be made available and, if so, which types of stakeholders should have access to it, there are various considerations, for example:

- The release of certain types of information may conflict with data privacy concerns, commercial or public policy interests—such as the promotion of innovation through appropriate intellectual property protections—and security interests, e.g., concerns about gaming and adversarial attacks. At the same time, such competing interests should not be permitted to be used, without specific justification, as a blanket cover for not adhering to due process, transparency, or accountability standards. The tension between these interests is particularly acute in the case of A/IS applied in a legal system, where the dignity, security, and liberty of individuals are at stake.[126]
- There is tension between the specific goal of explainability, which may argue for limits on system complexity, and system performance, which may be served by greater complexity, to the detriment of explainability.[127]
- One must carefully consider the question that is being asked in an inquiry into A/IS and what information transparency can actually produce to answer that question. Disclosure of A/IS algorithms or training data is, itself, insufficient to enable an auditor to determine whether the system was effective in a specific circumstance.[128] By analogy, transparency into drug manufacturing processes does not, itself, provide information about the actual effectiveness of a drug. Clinical trials provide that insight. In a legal system, an excessive focus on transparency-related information-gathering and assessment may overwhelm courts, legal practitioners, and law enforcement agencies. Meanwhile, other factors, such as measurement of effectiveness or operator competence, coupled with information on training data, may often suffice to ensure that there is a well-informed basis for trusting A/IS in a given circumstance.[129]

Given these competing considerations, arriving at a balance that is optimal for the functioning of a legal system and that has legitimacy in the eyes of the public will require an inclusive dialogue, bringing together the perspectives of those with an immediate stake in the proper functioning of a given technology, including those engaged in the design, development, procurement, deployment, operation, and validation of effectiveness of the technology, as well as those directly affected by the results of the technology; the perspectives of communities that may be indirectly impacted by the technology; and the perspectives of those with specialized expertise in ethics, government, and the law, such as jurists, regulators, and scholars. How the competing considerations should be balanced will also vary from one circumstance to another. Rather than aiming for universal transparency standards that would be applicable to all uses of A/IS within a legal system, transparency standards should allow for circumstance-dependent flexibility, in the context of the four constitutive components of trust discussed in this section.

Law

Additional note 1. The goals of transparency, e.g., answering a question as to why A/IS reached a given decision, may, in some cases, be better served by modes of explanation that do not involve examining an algorithm's terms or opening the "black box". A counterfactual explanation taking the form of, for example, "You were denied a loan because your annual income was £30,000; if your income had been £45,000, you would have been offered a loan," may provide more insight sooner than the disclosure of an algorithm.[130]

Additional note 2. The transparency principle intersects with other principles focused on fostering trust. More specifically, we note the following:

- **Transparency and effectiveness.** Information about the measurement of effectiveness can foster trust only if it is disclosed, i.e., only if there is transparency pertaining to the procedures and results of a measurement exercise.
- **Transparency and competence.** Transparency is essential in ensuring that the competencies required by the human operators of A/IS are known and met. At the same time, questions addressed by transparency extend beyond competence, while the questions addressed by competence extend beyond those answered by transparency.
- **Transparency and accountability.** Transparency is essential in determining accountability, but transparency serves purposes beyond accountability, while accountability seeks to answer questions not addressed directly by transparency.

Illustration—Transparency

In 2004, the city of Memphis, Tennessee, was experiencing an increase in crime rates that exceeded the national average. In response, in 2005, the city piloted a predictive policing program known as Blue CRUSH (Crime Reduction Utilizing Statistical History).[131] Blue CRUSH, developed in conjunction with the University of Memphis,[132] utilizes IBM's SPSS predictive analytics software to identify "hot spots": locations and times in which a given type of crime has a greater than average likelihood of occurring. The system generates its results through the analysis of a range of both historical data (type of crime, location, time of day, day of week, characteristics of victim, etc.) and live data provided by units on patrol. Equipped with the predictive crime map generated by the system, the Memphis Police Department can allocate resources dynamically to preempt or interrupt the target criminal activity. The precise response the department takes will vary with circumstance: deployment of a visible patrol car, deployment of an unmarked observer car, increasing vehicle stops in the area, undercover infiltration of the location, and so on.

The pilot program of Blue CRUSH focused on gang-related gun violence, which had been on the rise in Memphis prior to the pilot. The program showed an improvement, relative to incumbent methods, in the interdiction of such violence. Based on the success of the pilot, the scope of program was expanded, in 2007, for use throughout the city. By 2013, the policing efforts enabled by Blue CRUSH had helped to reduce overall crime in the city by over 30% and violent crime by 20%.[133] The program

Law

also enabled a dramatic increase in the rate at which crimes were solved: for cases handled by the department's Felony Assault Unit, the percentage of cases solved increased from 16% to nearly 70%.[134] And the program was cost effective: an analysis by Nucleus Research found that the program, when compared to the resources required to achieve the same results by traditional means, realized an annual benefit of approximately $7.2 million at a cost of just under $400,000.[135]

The story of the deployment of Blue CRUSH in the metropolitan Memphis area is not just about the technology; it is equally about the police personnel utilizing the technology and about the communities in which the technology was deployed. As noted by former Memphis Police Department Director Larry Godwin: "You can have all the technology in the world but you've got to have leadership, you've got to have accountability, you've got to have boots on the streets for it to succeed."[136] Crucial to the program's success was public support. Blue CRUSH represents a variety of predictive policing technology that limits itself to identifying the "where", the "when", and the "what" of criminal activity; it does not attempt to identify the "who", and therefore avoids a number of the privacy questions raised by technologies that do attempt to identify individual perpetrators. The technology will still, however, prompt responses by the police that could include more intrusive police activity in identified hot spots. The public must be willing to accept that activity, and that acceptance is won by transparency. To that end, Godwin and Janikowski held more than 200 community and neighborhood watch meetings to inform the public about the technology and how it would be used in policing their communities.[137] Without that level of transparency, it is doubtful that Blue CRUSH would have had the public support needed for its successful deployment.

Holding community meetings is an important step in building trust in a predictive policing program. As such programs become more widely implemented, however, and become more widely studied, trust may require more than town-hall meetings. Research into the programs has raised serious concerns about the ways in which they are implemented and their potential for perpetuating or even exacerbating historical bias.[138] Addressing these concerns will require more sophisticated and intrusive oversight than can be realized through community meetings.

Included among the questions that must be addressed are the following.

- In identifying hot spots, does the program rely primarily on arrest rates, which reflect (potentially biased) police activity, or does it rely on actual crime rates?
- What are the specific criteria for identifying a hot spot and are those criteria free of bias?[139]
- How accessible are the input data used to identify hot spots? Are they open to analysis by an independent expert?
- What mechanisms for oversight, review, and remediation of the program have been put in place? Such oversight should have access to the data used to train the system, the models used to identify hot spots, tests of the

Law

effectiveness of the system, and steps taken to remediate errors (such as bias) when they are uncovered.

As the public becomes more aware of the potential negative impact[140] of predictive policing programs, law enforcement agencies hoping to build trust in such programs will have to put in place transparency mechanisms that go beyond town-hall meetings and that enable a sophisticated response to such questions.

Recommendations

1. Governments and professional associations should facilitate dialogue among stakeholders—those engaged in the design, development, procurement, deployment, operation, and validation of effectiveness of the technology; those who may be immediately affected by the results of the technology; those who may be indirectly affected by the results of the technology, including the general public; and those with specialized expertise in ethics, politics, and the law—on the question of achieving a balance between transparency and other priorities, e.g., security, privacy, appropriate property rights, efficient and uniform response by the legal system, and more. In developing frameworks for achieving such balance, policymakers and professional associations should make allowance for circumstantial variation in how competing interests may be reconciled.
2. Policymakers developing frameworks for realizing transparency in A/IS applied to legal tasks should require that any frameworks they develop are sensitive both to the distinctions among the types of information that might be disclosed and to the distinctions among categories of individuals who may seek information about the design, operation, and results of a given system.
3. Policymakers developing frameworks for realizing transparency in A/IS to be adopted in a legal system should consider the role of appropriate protection for intellectual property, but should not allow those concerns to be used as a shield to prevent duly limited disclosure of information needed to ascertain whether A/IS meet acceptable standards of effectiveness, fairness, and safety. In developing such frameworks, policymakers should make allowance that the level of disclosure warranted will be, to some extent, dependent on what is at stake in a given circumstance.
4. Policymakers developing frameworks for realizing transparency in A/IS to be adopted in a legal system should consider the option of creating a role for a specially designated "public interest steward", or "trusted third party", who would be given access to sensitive information not accessible to others. Such a public interest steward would be charged with assessing the information to answer the public interest questions at hand but would be under obligation not to disclose the specifics of the information accessed in arriving at those answers.
5. Designers of A/IS should design their systems with a view to meeting transparency requirements, i.e., so as to enable some

categories of information about the system and its performance to be disclosed while enabling other categories, such as intellectual property, to be protected.

6. When negotiating contracts for the provision of A/IS products and services for use in the legal system, providers and buyers of A/IS should include contractual terms specifying what categories of information will be accessible to what categories of individuals who may seek information about the design, operation, and results of the A/IS.
7. In developing frameworks for realizing transparency in A/IS to be adopted in a legal system, policymakers should recognize that the information provided by other types of inquiries, e.g., examination of evidence of effectiveness or of operator competence, may in certain circumstances provide a more efficient means to informed trust in the effectiveness, fairness, and safety of the A/IS in question.
8. Governments should, where appropriate, work together with A/IS developers, as well as other stakeholders in the effective functioning of the legal system, to facilitate the creation of error-sharing mechanisms to enable the more effective identification, isolation, and correction of flaws in broadly deployed A/IS in their legal systems, such as a systematic facial recognition error in policing applications or in risk assessment algorithms. In developing such mechanisms, the question of precisely what information gets shared with precisely which groups may vary from application to application. All government efforts in this regard should be transparent and open to public scrutiny.
9. Governments should provide whistleblower protections to individuals who volunteer to offer information in situations where A/IS are not designed as claimed or operated as intended, or when their results are not interpreted correctly. For example, if a law enforcement agency is using facial recognition technology for a purpose that is illegal or unethical, or in a manner other than that in which it is intended to be used, an individual reporting that misuse should be given protection against reprisal. All government efforts in this regard should be transparent and open to public scrutiny.
10. Recommendation 1 under Issue 2, with respect to transparency.
11. Recommendation 2 under Issue 2, with respect to transparency.

Further Resources

- J. A. Kroll, J. Huey, S. Barocas, E. W. Felten, J. R. Reidenberg, D. G. Robinson, and H. Yu, "Accountable Algorithms," University of Pennsylvania Law Review, vol. 165, Feb. 2017.
- J. A. Kroll, "The fallacy of inscrutability," Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences, vol. 376, no. 2133, Oct. 2018.
- W. L. Perry, B. McInnis, C. C. Price, S. C. Smith, and J. S. Hollywood, "Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations," The RAND Corporation, 2013.
- A. D. Selbst and S. Barocas, "The Intuitive Appeal of Explainable Machines," Fordham Law Review, vol. 87, no. 3, 2018.

Law

- S. Wachter, B. Mittelstadt, and L. Floridi, "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation," International Data Privacy Law, vol. 7, no. 2, pp. 76-99, June 2017.
- S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR," Harvard Journal of Law & Technology, vol. 31, no. 2, 2018.
- R. Wexler, "Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System," Stanford Law Review, vol. 70, no. 5, pp. 1342-1429, 2017.

Section 2: Legal Status of A/IS

There has been much discussion about how to legally regulate A/IS-related technologies and the appropriate legal treatment of systems that deploy these technologies. Already, some lawmakers are wrestling with the issue of what status to apply to A/IS. Legal "personhood"—applied to humans and certain types of human organizations—is one possible option for framing such legal treatment, but granting that status to A/IS applications raises issues in multiple domains of human interaction.

Issue

What type of legal status (or other legal analytical framework) is appropriate for A/IS given (i) the legal issues raised by deployment of such technologies, and (ii) the desire to maximize the benefits of A/IS and minimize negative externalities?

Background

The convergence of A/IS and robotics technologies has led to the development of systems and devices resembling those of human beings in terms of their autonomy, ability to perform intellectual tasks, and, in the case of some robots, their physical appearance. As some types of A/IS begin to display characteristics resembling those of human actors, some governmental entities and private commentators have concluded that it is time to examine how legal regimes should categorize and treat various types of A/IS, often with an eye toward according A/IS a legal status beyond that of mere property. These entities have posited questions such as whether the law should treat such systems as legal persons.[141]

While legal personhood is a multifaceted concept, the essential feature of "full" legal personhood is the ability to participate autonomously within the legal system by having the right to sue and the capacity to be sued in court.[142] This allows legal "persons" to enter legally binding agreements, take independent action to enforce their own rights, and be held responsible for violations of the rights of others.

Conferring such status on A/IS seems initially remarkable until consideration is given to the long-standing legal personhood status granted to corporations, governmental entities, and the like—none of which are themselves human. Unlike these familiar legal entities, however, A/IS are not composed of—or necessarily controlled by—human beings. Recognizing A/IS as independent legal entities could therefore lead to abuses of that status, possibly by A/IS

Law

and certainly by the humans and legal entities who create or operate them, just as human shareholders and agents have abused the corporate form.[143] A/IS personhood is a significant departure from the legal traditions of both common law and civil law.[144]

Current legal frameworks provide a number of categories of legal status, other than full legal personhood, that could be used as analogues for the legal treatment of A/IS and how to allocate legal responsibility for harm caused by A/IS. At one extreme, legal systems could treat A/IS as mere products, tools, or other form of personal or intellectual property, and therefore subject to the applicable regimes of property law. Such treatment would have the benefit of simplifying allocation of responsibility for harm. It would, however, not account for the fact that A/IS, unlike other forms of property, may be capable of making legally significant decisions autonomously. In addition, if A/IS are to be treated as a form of property, governments and courts would have to establish rules regarding ownership, possession, and use by third parties. Other legal analogues may include the treatment of pets, livestock, wild animals, children, prisoners, and the legal principles of agency, guardianship, and powers of attorney.[145] Or perhaps A/IS are something entirely without precedent, raising the question of whether one or more types of A/IS might be assigned a hybrid, intermediate, or novel type of legal status?

Clarifying the legal status of A/IS in one or more jurisdictions is essential in removing the uncertainty associated with the obligations and expectations for organization and operation of these systems. Clarification along these lines will encourage more certain development and deployment of A/IS and will help clarify lines of legal responsibility and liability when A/IS cause harm. One of the problems of exploiting the existing status of legal personhood is that international treaties may bind multiple countries to follow the lead of a single legislature, as in the EU, making it impossible for a single country to experiment with the legal and economic consequences of such a strategy.

Recognizing A/IS as independent legal persons would limit or eliminate some human responsibility for subsequent decisions made by such A/IS. For example, under a theory of intervening causation, a hammer manufacturer is not held responsible when a burglar uses a hammer to break the window of a house. However, if similar "relief" from responsibility was available to the designers, developers, and users of A/IS, it will potentially reduce their incentives to ensure the safety of A/IS they design and use. In this example, legal issues that are applied in similar chain of causation settings—such as foreseeability, complicity, reasonable care, strict liability for unreasonably dangerous goods, and other precedential notions—will factor into the design process. Different jurisdictions may reach different conclusions about the nature of such causation chains, inviting future creative legal planners to consider how and where to pursue design, development, and deployment of future A/IS in order to receive the most beneficial legal treatment.

The legal status of A/IS thus intertwines with broader legal questions regarding how to ensure

Law

accountability and assign and allocate liability when A/IS cause harm. The question of legal personhood for A/IS, in particular, also interacts with broader ethical and practical questions on the extent to which A/IS should be treated as moral agents independent from their human designers and operators, whether recognition of A/IS personhood would enhance or detract from the purposes for which humans created the A/IS in the first place, and whether A/IS personhood facilitates of debilitates the widespread benefits of A/IS.

Some assert that because A/IS are at a very early stage of development, it is premature to choose a particular legal status or presumption in the many forms and settings in which those systems are and will be deployed. However, thoughtfully establishing a legal status early in the development could also provide crucial guidance to researchers, programmers, and developers. This uncertainty about legal status, coupled with the fact that multiple legal jurisdictions are already deploying A/IS—and each of them, as a sovereign entity, can regulate A/IS as it sees fit—suggests that there are multiple general frameworks that can and should be considered when assessing the legal status of A/IS.

Recommendations

1. While conferring full legal personhood on A/IS might bring some economic benefits, the technology has not yet developed to the point where it would be legally or morally appropriate to generally accord A/IS the rights and responsibilities inherent in the legal definition of personhood as it is now defined. Therefore, even absent the consideration of any negative ramifications from personhood status, it would be unwise to accord such status to A/IS at this time.
2. In determining what legal status, including granting A/IS legal rights short of full legal personhood, to accord to A/IS, government and industry stakeholders alike should: (1) identify the types of decisions and operations that should never be delegated to A/IS; and (2) determine what rules and standards will most effectively ensure human control over those decisions.
3. Governments and courts should review various potential legal models—including agency, animal law, and the other analogues discussed in this section—and assess whether they could serve as a proper basis for assigning and apportioning legal rights and responsibilities with respect to the deployment and use of A/IS.
4. In addition, governments should scrutinize existing laws—especially those governing business organizations—for mechanisms that could allow A/IS to have legal autonomy. If ambiguities or loopholes create a legal method for recognizing A/IS personhood, the government should review and, if appropriate, amend the pertinent laws.
5. Manufacturers and operators should learn how each jurisdiction would categorize a given autonomous and/or intelligent system and how each jurisdiction would treat harm caused by the system. Manufacturers and operators should be required to comply with the applicable laws of all jurisdictions in

Law

which that system could operate. In addition, manufacturers and operators should be aware of standards of performance and measurement promulgated by standards development organizations and agencies.

6. Stakeholders should be attentive to future developments that could warrant reconsideration of the legal status of A/IS. For example, if A/IS were developed that displayed self-awareness and consciousness, it may be appropriate to revisit the issue of whether they deserve a legal status on par with humans. Likewise, if legal systems underwent radical changes such that human rights and dignity no longer represented the primary guiding principle, the concept of full personhood for artificial entities may not represent the radical departure it might today. If the development of A/IS were to go in the opposite direction, and mechanisms were introduced allowing humans to control and predict the actions of A/IS easily and reliably, then the dangers of A/IS personhood would not be any greater than for well-established legal entities, such as corporations.
7. In considering whether to accord or expand legal protections, rights, and responsibilities to A/IS, governments should exercise utmost caution. Before according full legal personhood or a comparable legal status on A/IS, governments and courts should carefully consider whether doing so might limit how widely spread the benefits of A/IS are or could be, as well as whether doing so would harm human dignity and uniqueness of human identity. Governments and decision-makers at every level must work closely with regulators, representatives of civil society, industry actors, and other stakeholders to ensure that the interest of humanity—and not the interests of the autonomous systems themselves—remains the guiding principle.

Further Resources

- S. Bayern. "The Implications of Modern Business-Entity Law for the Regulation of Autonomous Systems." Stanford Technology Law Review 19, no. 1, pp. 93-112, 2015.
- S. Bayern, et al., "Company Law and Autonomous Systems: A Blueprint for Lawyers, Entrepreneurs, and Regulators." Hastings Science and Technology Law Journal, vol. 9, no. 2, pp. 135-162, 2017.
- D. Bhattacharyya. "Being, River: The Law, the Person and the Unthinkable." Humanities and Social Sciences Online, April 26, 2017.
- B. A. Garner, Black's Law Dictionary, 10th Edition, Thomas West, 2014.
- J. Bryson, et al., "Of, for, and by the people: the legal lacuna of synthetic persons," Artificial Intelligence Law 25, pp. 273-91, 2017.
- D. J. Calverley, "Android Science and Animal Rights, Does an Analogy Exist? " Connection Science 18, no. 4, pp. 403-417, 2006.
- D. J. Calverley, "Imagining a Non-Biological Machine as a Legal Person." AI & Society 22, pp. 403-417, 2008.
- R. Chatila, "Inclusion of Humanoid Robots in Human Society: Ethical Issues," in Springer Humanoid Robotics: A Reference, A. Goswami and P. Vadakkepat, Eds., Springer 2018.

Law

- European Parliament Resolution of 16 February 2017 (2015/2103(INL) with recommendations to the Commission on Civil Law Rules on Robotics, 2017.
- L. M. LoPucki, "Algorithmic Entities", 95 Washington University Law Review 887, 2018.
- J. S. Nelson, "Paper Dragon Thieves." Georgetown Law Journal 105, pp. 871-941, 2017.
- M. U. Scherer, "Of Wild Beasts and Digital Analogues: The Legal Status of Autonomous Systems." Nevada Law Journal 19, forthcoming 2018.
- M. U. Scherer, "Is Legal Personhood for AI Already Possible Under Current United States Laws?" Law and AI, May 14, 2017.
- L. B. Solum. "Legal Personhood for Artificial Intelligences." North Carolina Law Review 70, no. 4, pp. 1231–1287, 1992.
- J. F. Weaver. Robots Are People Too: How Siri, Google Car, and Artificial Intelligence Will Force Us to Change Our Laws. Santa Barbara, CA: Praeger, 2013.
- L. Zyga. "Incident of drunk man kicking humanoid robot raises legal questions," Techxplore, October 2, 2015.

Thanks to the Contributors

We wish to acknowledge all of the people who contributed to this chapter.

The Law Committee

- **John Casey** (Co-Chair) – Attorney-at-Law, Corporate, Wilson Sonsini Goodrich & Rosati, P.C.
- **Nicolas Economou** (Co-Chair) – Chief Executive Officer, H5; Chair, Science, Law and Society Initiative at The Future Society; Chair, Law Committee, Global Governance of AI Roundtable; Member, Council on Extended Intelligence
- **Aden Allen** – Senior Associate, Patent Litigation, Wilson Sonsini Goodrich & Rosati, P.C.
- **Miles Brundage** – Research Scientist (Policy), OpenAI; Research Associate, Future of Humanity Institute, University of Oxford; PhD candidate, Human and Social Dimensions of Science and Technology, Arizona State University
- **Thomas Burri** – Assistant Professor of International Law and European Law, University of St. Gallen (HSG), Switzerland
- **Ryan Calo** – Assistant Professor of Law, the School of Law at the University of Washington
- **Clemens Canel** – Referendar (Trainee Lawyer) at Hanseatisches Oberlandesgericht, graduate of the University of Texas School of Law and Bucerius Law School
- **Chandramauli Chaudhuri** – Senior Data Scientist; Fractal Analytics
- **Danielle Keats Citron** – Lois K. Macht Research Professor & Professor of Law, University of Maryland Carey School of Law
- **Fernando Delgado** – PhD Student, Information Science, Cornell University.
- **Deven Desai** – Associate Professor of Law and Ethics, Georgia Institute of Technology, Scheller College of Business
- **Julien Durand** – International Technology Lawyer; Executive Director Compliance & Ethics, Amgen Biotechnology
- **Todd Elmer**, JD – Member of the Board of Directors, National Science and Technology Medals Foundation
- **Kay Firth-Butterfield** – Project Head, AI and Machine Learning at the World Economic Forum. Founding Advocate of AI-Global; Senior Fellow and Distinguished Scholar, Robert S. Strauss Center for International Security and Law, University of Texas, Austin; Co-Founder, Consortium for Law and Ethics of Artificial Intelligence and Robotics, University of Texas, Austin; Partner, Cognitive Finance Group, London, U.K.
- **Tom D. Grant** – Fellow, Wolfson College; Senior Associate of the Lauterpacht Centre for International Law, University of Cambridge, U.K.

Law

- **Cordel Green** – Attorney-at-Law; Executive Director, Broadcasting Commission—Jamaica
- **Maura R. Grossman** – Research Professor, David R. Cheriton School of Computer Science, University of Waterloo; Adjunct Professor, Osgoode Hall Law School, York University
- **Bruce Hedin** – Principal Scientist, H5
- **Daniel Hinkle** – Senior State Affairs Counsel for the American Association for Justice
- **Derek Jinks** – Marrs McLean Professor in Law, University of Texas Law School; Director, Consortium on Law and Ethics of Artificial Intelligence and Robotics (CLEAR), Robert S. Strauss Center for International Security and Law, University of Texas.
- **Nicolas Jupillat** – Adjunct Professor, University of Detroit Mercy School of Law
- **Marwan Kawadri** – Analyst, Founders Intelligence; Research Associate, The Future Society.
- **Mauricio K. Kimura** – Lawyer; PhD student, Faculty of Law, University of Waikato, New Zealand; LLM from George Washington University, Washington DC, USA; Bachelor of Laws from Sao Bernardo do Campo School of Law, Brazil
- **Irene Kitsara** – Lawyer; IP Information Officer, Access to Information and Knowledge Division, World Intellectual Property Organization, Switzerland
- **Timothy Lau**, **J.D., Sc.D.** – Research Associate, Federal Judicial Center
- **Mark Lyon** – Attorney-at-Law, Chair, Artificial Intelligence and Autonomous Systems Practice Group at Gibson, Dunn & Crutcher LLP
- **Gary Marchant** – Regents' Professor of Law, Lincoln Professor of Emerging Technologies, Law and Ethics, Arizona State University
- **Nicolas Miailhe** – Co-Founder & President, The Future Society; Member, AI Expert Group at the OECD; Member, Global Council on Extended Intelligence; Senior Visiting Research Fellow, Program on Science Technology and Society at Harvard Kennedy School. Lecturer, Paris School of International Affairs (Sciences Po); Visiting Professor, IE School of Global and Public Affairs
- **Paul Moseley** – Master's student, Electrical Engineering, Southern Methodist University; graduate of the University of Texas School of Law
- **Florian Ostmann** – Policy Fellow, The Alan Turing Institute
- **Pedro Pavón** – Assistant General Counsel, Global Data Protection, Honeywell
- **Josephine Png** – AI Policy Researcher and Deputy Project Manager, The Future Society; budding barrister; and BA Chinese and Law, School of Oriental and African Studies
- **Matthew Scherer** – Attorney at Littler Mendelson, P.C., and legal scholar based in Portland, Oregon, USA; Editor, LawAndAI.com
- **Bardo Schettini Gherardini** – Independent Legal Advisor on standardization, AI and robotics

Law

- **Jason Tashea** – Founder, Justice Codes and adjunct law professor at Georgetown Law Center
- **Yan Tougas** – Global Ethics & Compliance Officer, United Technologies Corporation; Adjunct Professor, Law & Ethics, University of Connecticut School of Business; Fellow, Ethics & Compliance Initiative; Kallman Executive Fellow, Bentley University Hoffman Center for Business Ethics
- **Sandra Wachter** – Lawyer and Research Fellow in Data Ethics, AI and Robotics, Oxford Internet Institute, University of Oxford
- **Axel Walz** – Lawyer; Senior Research Fellow at the Max Planck Institute for Innovation and Competition, Germany. (Member until October 31, 2018)
- **John Frank Weaver** – Lawyer, McLane Middleton, P.A; Columnist for and Member of Board of Editors of *Journal of Robotics, Artificial Intelligence & Law*; Contributing Writer for *Slate*; Author, *Robots Are People Too*
- **Julius Weitzdörfer** – Affiliated Lecturer, Faculty of Law, University of Cambridge; Research Associate, Centre for the Study of Existential Risk, University of Cambridge
- **Yueh-Hsuan Weng** – Assistant Professor, Frontier Research Institute for Interdisciplinary Sciences (FRIS), Tohoku University; Fellow, Transatlantic Technology Law Forum (TTLF), Stanford Law School
- **Andrew Woods** – Associate Professor of Law, University of Arizona

For a full listing of all IEEE Global Initiative Members, visit standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ec_bios.pdf.

For information on disclaimers associated with EAD1e, see *How the Document Was Prepared*.

Law

The Law Committee of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems would like to thank the following individuals for taking the time to offer valuable feedback and suggestions on Section 1 of the Law Chapter, "Norms for the Trustworthy Adoption of A/IS in Legal Systems". Each of these contributors offered comments in an individual capacity, not in the name of the organization for which they work. The final version of the Section does not necessarily incorporate all comments or reflect the views of each contributor.

- **Rediet Abebe**, PhD Candidate, Department of Computer Science, Cornell University; cofounder, Mechanism Design for Social Good; cofounder, Black in AI.
- **Ifeoma Ajunwa**, Assistant Professor, Labor & Employment Law, Cornell Industrial and Labor Relations School; faculty Associate at Harvard Law, Berkman Klein Center.
- **Jason R. Baron**, of counsel, Drinker Biddle; co-chair, Information Governance Initiative; former Director of Litigation, United States National Archives and Records Administration.
- **Irakli Beridze**, Head, Centre for Artificial Intelligence and Robotics, United Nations (UNICRI).
- **Juan Carlos Botero**, Law Professor, Pontificia Universidad Javeriana, Bogota; former Executive Director, World Justice Project.
- **Anne Carblanc**, Principal Administrator, Information, Communications and Consumer Policy (ICCP) Division, Directorate for Science, Technology and Industry, OECD; former criminal investigations judge (juge d'instruction), Tribunal of Paris.
- **Gallia Daor**, Policy Analyst, OECD.
- **Lydia de la Torre**, Privacy Law Fellow, Santa Clara University.
- **Isabela Ferrari**, Federal Judge, Federal Court, Rio de Janeiro, Brazil.
- **Albert Fox Cahn**, Founder and Executive Director, Surveillance Technology Oversight Project; former Legal Director, CAIR-NY.
- **Paul W. Grimm**, United States District Judge, United States District Court for the District of Maryland.
- **Gillian Hadfield**, Professor of Law and Professor of Strategic Management, University of Toronto; Member, World Economic Forum Future Council for Agile Governance.
- **Sheila Jasanoff**, Pforzheimer Professor of Science and Technology Studies, Harvard Kennedy School of Government.
- **Baroness Beeban Kidron**, OBE, Member, United Kingdom House of Lords.
- **Eva Kaili**, Member, European Parliament; Chair, European Parliament Science and Technology Options Assessment body (STOA).
- **Mantalena Kaili**, cofounder, European Law Observatory on New Technologies.
- **Jon Kleinberg**, Tisch University Professor, Departments of Computer Science and Information Science, Cornell University; member of the National Academy of Sciences, the National Academy of Engineering, and the American Academy of Arts and Sciences.
- **Shuang Lu Frost**, Teaching Fellow, PhD candidate, Department of Anthropology, Harvard University.

Law

- **Arthur R. Miller CBE**, University Professor, New York University; former Bruce Bromley Professor of Law, Harvard Law School.
- **Manuel Muñiz**, Dean and Rafael del Pino Professor of Practice of Global Leadership, IE School of Global and Public Affairs, Madrid; Senior Associate, Belfer Center, Harvard University.
- **Erik Navarro Wolkart**, Federal Judge, Federal Court, Rio de Janeiro, Brazil.
- **Aileen Nielsen**, chair, Science and Law Committee, New York City Bar Association.
- **Michael Philips**, Assistant General Counsel, Microsoft.
- **Dinah PoKempner**, General Counsel, Human Rights Watch.
- **Irina Raicu**, Director, Internet Ethics Program, Markkula Center for Applied Ethics, Santa Clara University.
- **David Robinson**, Visiting Scientist, AI Policy and Practice Initiative, Cornell University; Adjunct Professor of Law, Georgetown University Law Center; Managing Director (on leave), Upturn.
- **Alanna Rutherford**, Vice President, Global Litigation & Competition, Visa.
- **George Socha, Esq.**, Consulting Managing Director, BDO USA; co-founder, Electronic Discovery Reference Model (EDRM) and Information Governance Reference Model (IGRM).
- **Lee Tiedrich**, Partner, IP/Technology Transactions, and Co-Chair, Artificial Intelligence Initiative, Covington & Burling LLP.
- **Darrell M. West**, VP, Governance Studies, Director, Center for Technology Innovation, Douglas Dillon Chair in Governance Studies, Brookings Institution.
- **Bendert Zevenbergen**, Research Fellow, Center for Information Technology Policy, Princeton University; Researcher, Oxford Internet Institute.
- **Jiyu Zhang**, Associate Professor and Executive Director of the Law and Technology Institute, Renmin University of China School of Law.
- **Peter Zimroth**, Director, New York University Center on Civil Justice; retired partner, Arnold & Porter; former Assistant US Attorney, Southern District of New York.

Endnotes

[1] See S. Jasanoff, "Governing Innovation: The Social Contract and the Democratic Imagination," Seminar, vol. 597, pp. 16-25, May 2009.

[2] As articulated in *EAD* General Principles 1 (Human Rights), 2 (Well-Being), and 3 (Data Agency). See also *EAD* Chapter, "Classical Ethics in A/IS," In applying A/IS in pursuit of these goals, tradeoffs are inevitable. Some applications of predictive policing, for example, may reduce crime, and so enhance well-being, but may do so at the cost of impinging on a right to privacy or weakening protections against unwarranted search and seizure. How these tradeoffs are negotiated may vary with cultural and legal traditions.

[3] Risks and benefits, and their perception, are neither always well-defined at the outset nor static over time. Social expectations and even ideas of lawfulness constantly evolve. For example, if younger generations, accustomed to the use of social networking technologies, have lower expectations of privacy than older generations, should this be deemed to be a benefit to society, a risk, or neither?

[4] Regarding the nature of the guidance provided in this section: Artificial intelligence, like many other domains relied on by the legal realm (e.g., medical and accounting forensics, ballistics, or economic analysis), is a scientific discipline distinct from the law. Its effective and safe design and operation have underpinnings in academic and professional competencies in computer science, linguistics, data science, statistics, and related technical fields. Lawyers, judges, and law enforcement officers increasingly draw on these fields, directly or indirectly, as A/IS are progressively adopted in the legal system. This document does not seek to offer legal advice to lawyers, courts, or law enforcement agencies on how to practice their professions or enforce the law in their jurisdictions around the globe. Instead, it seeks to help ensure that A/IS and their operators in a given legal system can be trusted by lawyers, courts, and law enforcement agencies, and civil society at large, to perform effectively and safely. Such effective and safe operation of A/IS holds the potential of producing substantial benefits for the legal system, while protecting all of its participants from the ethical, professional, and business risks, or personal jeopardy, that may result from the intentional, unintentional, uninformed, or incompetent procurement and operation of artificial intelligence.

[5] See Rensselaer Polytechnic Institute, "A Conversation with Chief Justice John G. Roberts, Jr.," April 11, 2017. YouTube video, 40:12. April 12, 2017. [Online]. Available: https://www.youtube.com/watch?v=TuZEKlRgDEg.

[6] "Uninformed avoidance of adoption" can be one of two types: (a) avoidance of adoption when the information needed to enable sound decisions is available but is not taken into

consideration, and (b) avoidance of adoption when the information needed to enable sound decisions is simply not available. Unlike the former type of avoidance, the latter type is a prudent and well-reasoned avoidance of adoption and, pending better information, is the course recommended by a number experts and nonexperts.

[7] For purposes of this chapter, we have made the deliberate choice to focus on these four principles without taking a prior position on where the deployment of A/IS may or may not be acceptable in legal systems. Where these principles cannot be adequately operationalized, it would follow that the deployment of A/IS in a legal system cannot be trusted. Where A/IS can be evidenced to meet desired thresholds for each duly operationalized principle, it would follow that their deployment can be trusted. Such information is intended to facilitate, not preempt, the indispensable public policy dialogue on the extent to which A/IS should be relied upon to meet the specific needs of the legal systems of societies around the world.

[8] It is beyond the scope of this chapter to discuss the process through which such adherence may become institutionalized in the complex legal, technological, political, and cultural dynamics in which sociotechnical innovation occurs. It is worth noting, however, that this process typically involves four steps. First, a wide range of market and culture-driven practices emerge. Second, a set of best practices arises, reflecting a group's willingness to adopt certain rules. Third, some of these best practices are formulated into standards, which enable enforcement (through private contracts, professional codes of practice, or legislation). Finally, those enforceable standards render the performance of some activities sufficiently reliable and predictable to enable trustworthy operation at the scale of society. Where these elements (rulemaking, enforcement, scalable operation) are present, new institutions are born.

[9] For a discussion of the definition of A/IS, see the Terminology Update in the Executive Summary of EAD. The principles outlined in this section as constitutive of "informed trust" do not depend on a precise, consensus definition of A/IS and are, in fact, designed to be enable successful operationalization under a broad range of definitions.

[10] Such as Gross Domestic Product (GDP), Gross National Income (GNI) per capita, the WEF Global Competitiveness Index, and others.

[11] Such as life expectancy, infant mortality rate, and literacy rate, as well as composite indices such as the Human Development Index, the Inequality-Adjusted Human Development Index, the OECD Framework for Measuring Well-being and Progress, and others. For more on measures of well-being, see the *EAD* chapter on "Well-being".

[12] See United Nation General Assembly, Universal Declaration of Human Rights, Dec. 10, 1948, available: http://www.un.org/en/universal-declaration-human-rights/index.html; see also United Nations Office of the High Commissioner: Human Rights, The Vienna Declaration and Programme of Action, June 25, 1993, available: https://www.ohchr.org/en/professionalinterest/pages/vienna.aspx.

Law

[13] See UNICEF, Convention on the Rights of the Child, Nov. 4, 2014, available: https://www.unicef.org/crc/index_30160.html.

[14] See United Nations Security Council, "The Rule of Law and Transitional Justice in Conflict and Post-conflict Societies: Report of the Secretary General," *Report S/2004/616* (2004).

[15] See The World Economic Forum, *The Global Competitiveness Report: 2018*, ed. K. Schwab (2018), pp. 12ff.

[16] See A. Brunetti, G. Kisunko, and B. Weder, "Credibility of Rules and Economic Growth: Evidence from a Worldwide Survey of the Private Sector," The World Bank Economic Review, vol. 12, no. 3, pp. 353–384, 1998. Available: https://doi.org/10.1093/wber/12.3.353; see also World Bank, World Development Report 2017: Governance and the Law, Jan. 2017. Available: doi.org/10.1596/978-1-4648-0950-7.

[17] The question of intellectual property law in an era of rapidly advancing technology (both A/IS and other technologies) is a complex and often contentious one involving legal, economic, and ethical considerations. We have not yet studied the question in sufficient depth to reach a consensus on the issues raised. We may examine the issues in depth in a future version of *EAD*. For a forum in which such issues are discussed, see the Berkeley-Stanford Advanced Patent Law Institute. See also The World Economic Forum, "Artificial Intelligence Collides with Patent Law." April 2018. Available: http://www3.weforum.org/docs/WEF_48540_WP_End_of_Innovation_Protecting_Patent_Law.pdf.

[18] A component of human dignity is privacy, and a component of privacy is protection and control of one's data; in this regard, frameworks such as the EU's General Data Protection Regulation (GDPR) and the Council of Europe's "Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data" have a role to play in setting standards for how legal systems can protect data privacy. See also *EAD* General Principle 3 (Data Agency).

[19] Frameworks such as the Universal Declaration of Human Rights and the Vienna Declaration and Programme of Action (VDPA) have a role to play in articulating human-rights standards to which legal systems should adhere. See also *EAD* General Principle 1 (Human Rights).

[20] For more on the importance of measures of well-being beyond GDP, see *EAD* General Principle 2 (Well-being).

[21] For a conceptual framework enabling the country-by-country assessment of the Rule of Law, see World Justice Project, *Rule of Law Index*. 2018. url: https://worldjusticeproject.org/sites/default/files/documents/WJP-ROLI-2018-June-Online-Edition_0.pdf.

[22] *See* D. Kennedy, "The 'Rule of Law,' Political Choices and Development Common Sense," in The New Law and Economic Development: A Critical Appraisal, D. M. Trubek and A. Santos, Ed. Cambridge: Cambridge University Press, 2006, pp. 156-157; see also A. Sen, *Development as Freedom*. New York: Alfred A. Knopf, 1999.

Law

[23] *See* Kennedy (2006): pp. 168-169. "The idea that building 'the rule of law' might *itself* be a development strategy encourages the hope that choosing law *in general* could substitute for all the perplexing political and economic choices that have been at the center of development policy making for half a century. The politics of allocation is submerged. Although a legal regime offers an arena to contest those choices, it cannot substitute for them."

[24] *Fairness* (as well as *bias*) can be defined in more than one way. For purposes of this chapter, a commitment is not made to any one definition—and indeed, it may not be either desirable or feasible to arrive at a single definition that would be applied in all circumstances. The trust principles proposed in the chapter (Effectiveness, Competence, Accountability, and Transparency) are defined such that they will provide information that will allow the testing of an application of A/IS against any fairness criteria.

[25] The confidentiality of jury deliberations, certain sensitive cases, and personal data are some of the considerations that influence the extent of appropriate public examination and oversight mechanisms.

[26] The avoidance of negative consequences is important to note in relation to effectiveness. The law can be used for malevolent or intensely disputed purposes (for example, the quashing of dissent or mass incarceration). The instruments of the law, including A/IS, can render the advancement of such purposes more effective to the detriment of democratic values, human rights, and human well-being.

[27] Studies conducted by the US National Institute of Standards and Technology (NIST) between 2006 and 2011, known as the US NIST Text REtrieval Conference (TREC) Legal Track, suggest that some A/IS-enabled processes, if operated by trained experts in the relevant scientific fields, can be more effective (or accurate) than human attorneys in correctly identifying case-relevant information in large data sets. NIST has a long-standing reputation for cultivating trust in technology by participating in the development of standards and metrics that strengthen measurement science and make technology more secure, usable, interoperable, and reliable. This work is critical in the A/IS space to ensure public trust of rapidly evolving technologies so that we can benefit from all that this field has to promise.

[28] In describing the potential A/IS have for aiding in the auditing of decisions made in the civil and criminal justice systems, we are envisioning them acting as aids to a competent human auditor (see Issue 4) in the context of internal or judicial review.

[29] Of course, the use of A/IS in improving the effectiveness of law enforcement may raise concerns about other aspects of well-being, such as privacy and the rise of the surveillance state, cf. Minority Report (2002). If A/IS are to be used for law enforcement, steps must be taken to ensure that they are used, and that citizens trust that they will be used, in ways that are conducive to ethical law enforcement and individual well-being (see Issue 2).

Law

[30] A/IS may also provide assistance in carrying out legal tasks associated with larger transactions, such as evaluating contracts for risk in connection with a M&A transaction or reporting exposure to regulators.

[31] The recommendations provided in this chapter (both under this issue and under the other issues discussed in the chapter) are intended to give general guidance as to how those with a stake in the just and effective operation of a legal system can develop norms for the trustworthy adoption of A/IS in the legal system. The specific ways in which the recommendations are operationalized will vary from society to society and from jurisdiction to jurisdiction.

[32] See "Global Governance of AI Roundtable: Summary Report 2018," World Government Summit, 2018: p. 32. Available: https://www.worldgovernmentsummit.org/api/publications/document?id=ff6c88c5-e97c-6578-b2f8-ff0000a7ddb6. (The February 2018 Dubai Global Governance of AI Roundtable brought together ninety leading thinkers on AI governance.)

[33] *See State v Loomis*, 881 N.W.2d 749 (Wis. 2016), *cert. denied* (2017); see also "Criminal Law—Sentencing Guidelines—Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing—State v. Loomis, 881 N.W.2d 749 (Wis. 2016)," Harvard Law Review, vol. 130, no. 5, pp. 1535-1536, 2017. Available: http://harvardlawreview.org/wp-content/uploads/2017/03/1530-1537_online.pdf; see also K. Freeman, "Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis," North Carolina Journal of Law and Technology, vol. 18, no. 5, pp. 75-76, 2016. Available: https://scholarship.law.unc.edu/ncjolt/vol18/iss5/3/.

[34] An example of an initiative that seeks to bridge the gap between technical and legal expertise is the Artificial Intelligence Legal Challenge, held at Ryerson University and sponsored by Canada's Ministry of the Attorney General: http://www.legalinnovationzone.ca/press_release/ryersons-legal-innovation-zone-announces-winners-of-ai-legal-challenge/.

[35] And, in addressing the challenges, consideration must be given to existing modes of proposing and approving innovation in the legal system. Trust in A/IS will be undermined if they are viewed as not having been vetted via established processes.

[36] For an overview of risk and risk management, see Working Party on Security and Privacy in the Digital Economy, Background Report for Ministerial Panel 3.2, Directorate for Science, Technology and Innovation, Committee on Digital Economy Policy, Managing Digital Security and Privacy Risk, OECD, June 1, 2016; see p. 5.

[37] It is worth emphasizing the "informed" qualifier we attach to trust here. Far from advocating for a "blind trust" in A/IS, we argue that A/IS should be adopted only when we have sound evidence of their effectiveness, when we can be confident of the competence of their operators, when we have assurances that these systems allow for the attribution of responsibility for outcomes (both positive and negative), and when we have clear views into their operation. Without those conditions, we would argue that A/IS *should not be adopted* in the legal system.

Law

[38] The importance of testing the effectiveness of advanced technologies, including A/IS, in the legal system (and beyond) is not new: it was highlighted by Judge Paul W. Grimm in an important early ruling on legal fact-finding, *Victor Stanley v. Creative Pipe, Inc.*, 250 F.R.D. 251, 257 (D. Md. 2008), followed, among others, by the influential research and educational institute The Sedona Conference as well as the International Organization for Standardization (ISO). See *An Open Letter to Law Firms and Companies in the Legal Tech Sector*, The Sedona Conference (2009), and *Commentary on Achieving Quality in the E-Discovery Process* (2013): 7; ISO standard on electronic discovery (ISO/IEC 27050-3:2017): 19. Most recently, in the summary report of the Global Governance of AI Roundtable at the 2018 World Government Summit, Omar bin Sultan Al Olama, Minister of State for Artificial Intelligence of the UAE, highlighted the importance of "empirical information" in assessing the suitability of A/IS.

[39] In the terminology of software development, *verification* is a demonstration that a given application meets a narrowly defined requirement; *validation* is a demonstration that the application answers its real-world use case. When we speak of gathering evidence of the effectiveness of A/IS, we are speaking of validation.

[40] Standards may include compliance with defined professional competence or other ethical requirements, but also other types of standards, such as data standards. Data standards may serve as "a digital lingua franca" with the potential of both supporting broad-based technological innovation (including A/IS innovation) in a legal system and facilitating access to justice. As part of interactive technology solutions, appropriate data standards may help connect the ordinary citizen to the appropriate resources and information for his or her legal needs. For a discussion of open data standards in the context of the US court system, see D. Colarusso and E. J. Rickard, "Speaking the Same Language: Data Standards and Disruptive Technologies in the Administration of Justice," Suffolk University Law Review, vol. L387, 2017.

[41] For measurement of bias in facial recognition software, see C. Garvie, A. M. Bedoya, and J. Frankle, "The Perpetual Line-Up: Unregulated Police Face Recognition in America," Georgetown Law, Center on Privacy & Technology, Oct. 2016. Available: https://www.perpetuallineup.org/.

[42] The inclusion of such collateral effects in assessing effectiveness is an important element in overcoming the apparent "black box" or inscrutable nature of A/IS. See, for example, J. A. Kroll, "The fallacy of inscrutability," Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences, vol. 376, no. 2133, Oct. 2018. Available: doi.org/0.1098/rsta.2018.0084. The study addresses, among other questions, "how measurement of a system beyond understanding of its internals and its design can help to defeat inscrutability."

[43] The question of the salience of collateral impact will vary with the specific application of A/IS. For example, false positives in document review related to fact-finding will generally not raise acute ethical issues, but false positives

in predictive policing or sentencing will. In these latter domains, complex and sometimes unsettled issues of fairness arise, particularly when social norms of fairness change regionally and over time (sometimes rapidly). Any A/IS that was designed to replicate some notion of fairness would need to demonstrate its effectiveness, first, at replicating prevailing notions of fairness that have legitimacy in society, and second, at responding to evolutions in such notions of fairness. In the current state of A/IS, in which no system has been able to demonstrate consistent effectiveness in either of the above regards, it is essential that great discretion be exercised in considering any reliance on A/IS in domains such as sentencing and predictive policing.

[44] These exercises go by various names in the literature: *effectiveness evaluations*, *benchmarking exercises*, *validation studies*, and so on. See, for example, the definition of *validation study* in AINOW's 2018 *Algorithmic Accountability Toolkit* (https://ainowinstitute.org/aap-toolkit.pdf), p. 29. For our purposes, what matters is that the exercise be one that collects, in a scientifically sound manner, evidence of how "fit for purpose" any given A/IS are.

[45] This feature of evaluation design is important, as only tasks that accurately reflect real-world conditions and objectives (which may include the avoidance of unintended consequences, such as racial bias) will provide compelling guidance as to the suitability of an application for adoption in the real world.

[46] For TREC generally, see: https://trec.nist.gov/. For the TREC Legal Track specifically, see: https://trec-legal.umiacs.umd.edu/.

[47] When a complex system can be broken down into separate component systems, it may be appropriate to assess either the effectiveness of each component, or that of the end-to-end application as a whole (including human operators), depending on the specific question to be answered.

[48] Qualitative considerations may also help counter attempts to "game the system" (i.e., attempts to use bad-faith methods to meet a specific numerical target); see B. Hedin, D. Brassil, and A. Jones, "On the Place of Measurement in E-Discovery," in Perspectives on Predictive Coding and Other Advanced Search Methods for the Legal Practitioner, ed. J. R. Baron, R. C. Losey, and M. D. Berman. Chicago: American Bar Association, 2016, p. 415f.

[49] Even in fact-finding, accurate extraction of facts does not eliminate the need for reasoned judgment as to the significance of the facts in the context of specific circumstances and cultural considerations. Used properly, A/IS will advance the spirit of the law, not just the letter of the law.

[50] Electronic discovery is the task of searching through large collections of electronically stored information (ESI) for material relevant to civil and criminal litigation and investigations. Among applications of A/IS to legal tasks and questions, the application to legal discovery is probably the most "mature," as measured against the criteria of having been tested, assessed and approved by courts, and adopted fairly widely across various jurisdictions.

Law

[51] While there is general consensus about the importance of these metrics in gauging effectiveness in legal discovery, there is not a consensus around the precise values for those metrics that must be met for a discovery effort to be acceptable. That is a good thing, as the precise value that should be attained, and demonstrated to have been attained, in any given matter will be dependent on, and proportional to, the specific facts and circumstances of that matter.

[52] Different domains of application of A/IS to legal matters will vary not only with regard to the availability of consensus metrics of effectiveness, but also with regard to conditions that affect the challenge of measuring effectiveness: availability of data, impact of social bias, and sensitivity to privacy concerns all affect how difficult it may be to arrive at consensus protocols for gauging effectiveness. In the case of defining an effectiveness metric for A/IS used in support of sentencing decisions, one challenge is that, while it is easy to find when an individual who has been released commits a crime (or is convicted of committing a crime), it is difficult to assess when an individual who was not released would have committed a crime. For a discussion of the challenges in measuring the effectiveness of tools designed to assess flight risk, see M. T. Stevenson, "Assessing Risk Assessment in Action." Minnesota Law Review, vol. 103, 2018. Available: doi.org/10.2139/ssrn.3016088.

[53] Sound measurement may also serve as an effective antidote to the unsubstantiated claims sometimes made regarding the effectiveness of certain applications of A/IS to legal matters (e.g., flight risk assessment technologies); see Stevenson, "Assessing Risk Assessment". Unsubstantiated claims are an appropriate source of an *informed distrust* in A/IS. Such well-founded distrust can be addressed only with truly meaningful and sound measures that provide accurate information regarding the capabilities and limitations of a given system.

[54] See the discussion under "Illustration—Effectiveness" in this chapter.

[55] For more on principles for data protection, see the EAD chapter "Personal Data and Individual Agency".

[56] The importance of validation by practitioners is reflected in The European Commission's High-Level Expert Group on Artificial Intelligence Draft Ethics Guidelines for Trustworthy AI: "Testing and validation of the system should thus occur as early as possible and be iterative, ensuring the system behaves as intended throughout its entire life cycle *and especially after deployment.*" (Emphasis added.) See High-Level Expert Group on Artificial Intelligence, "DRAFT Ethics Guidelines for Trustworthy AI: Working Document for Stakeholders' Consultation," The European Commission. Brussels, Belgium: Dec. 18, 2018.

[57] That scrutiny need not extend to IP or other protected information (e.g., attorney work product). Validation methods and results are a matter of numbers and procedures for obtaining the numbers, and their disclosure would not impinge on safeguards against the disclosure of legitimately protected information.

Law

[58] A recent matter from the US legal system illustrates how a failure to disclose the results of a validation exercise can limit the exercise's ability to achieve its intended purpose. In *Winfield v. City of New York* (Opinion & Order. 15-CV-05236 [LTS] [KHP]. SDNY 2017), a party had utilized the A/IS-enabled system to conduct a review of documents for relevance to the matter being litigated. When the accuracy and completeness of the results of that review were challenged by the requesting party, the producing party disclosed that it had, in fact, conducted validation of its results. Rather than requiring that the producing party simply disclose the results of the validation to the requesting party, the judge overseeing the dispute chose to review the results herself *in camera*, without providing access to the requesting party. Although the judge then said that the evidence she was provided supported the accuracy and completeness of the review, the requesting party could not itself examine either the evidence or the methods whereby it was obtained, and so could not gain confidence in the results. That confidence comes only from examining the metrics and the procedures followed in obtaining them. Moreover, the results of a validation exercise, which are usually simple numbers that reflect sampling procedures, can be disclosed without revealing the content of any documents, any proprietary tools or methods, or any attorney work product. If the purpose of conducting a validation exercise is to gather evidence of the effectiveness of a process, in the event that the process is challenged, keeping that evidence hidden from those who would challenge the process limits the ability of the validation exercise to achieve its intended purpose.

[59] https://www.nist.gov/.

[60] TREC Legal Track (2006-2011): https://trec-legal.umiacs.umd.edu/.

[61] The statistical evidence in question here is statistical evidence of the effectiveness of A/IS applied to the task of discovery; it is not statistical evidence of facts actually at issue in litigation. Courts may have different rules for the admissibility of the two kinds of statistical evidence (and there will be jurisdictional differences on these questions).

[62] It is important to underscore that, whereas developers and operators of A/IS should be able to derive sound measurements of effectiveness, the courts should determine what level of effectiveness—what score—should be demonstrated to have been achieved, based on the facts and circumstances of a given matter. In some instances, the cost (in terms of sample sizes, resources required to review the samples, and so on) of demonstrating the achievement of a high score will be disproportionate to the stakes of a given matter. In others, for example, a major securities fraud claim that potentially affects thousands of citizens, a court might justifiably demand a demonstration of the achievement of a very high score, irrespective of cost. Demonstrations of the effectiveness of A/IS (and of their operators) are instruments in support of, not in substitution of, judicial decision-making.

[63] See, for example, B. Hedin, S. Tomlinson, J. R. Baron, and D. W. Oard, "Overview of the TREC 2009 Legal Track," in *NIST Special Publication: SP 500-278, The Eighteenth Text REtrieval Conference (TREC 2009) Proceedings* (2009).

Law

[64] See M. R. Grossman and G. V. Cormack, "Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review," Richmond Journal of Law and Technology, vol. 17, no. 3, 2011. Available: http://jolt.richmond.edu/jolt-archive/v17i3/article11.pdf. Note that the two systems that conclusively demonstrated "better than human" performance took methodologically distinct approaches, but they shared the characteristic of having been designed, operated, and measured for accuracy by scientifically trained experts.

[65] *Da Silva Moore v. Publicis Groupe*, 2012 WL 607412 (S.D.N.Y. Feb. 24, 2012). See also A. Peck, "Search, Forward," Legaltech News. Oct. 1, 2011. Available: https://www.law.com/legaltechnews/almID/1202516530534Search-Forward/.

[66] The fact that NIST has as important role to play in developing standards for the measurement of the safety and security of A/IS was recognized in a recent (September, 2018) report from the U.S. House of Representatives: "At minimum, a widely agreed upon standard for measuring the safety and security of AI products and applications should precede any new regulations. ... The National Institute of Standards and Technology (NIST) is situated to be a key player in developing standards." (Will Hurd and Robin Kelly, "Rise of the Machines: Artificial Intelligence and its Growing Impact on U.S. Policy," U.S. House of Representatives—Committee on Oversight and Government Reform—Subcommittee on Information Technology, September, 2018).

[67] The competence principle is intended to apply to the post design operation of A/IS. Of course, that does not mean that designers and developers of A/IS are free of responsibility for their systems' outcomes. As discussed in the background to this issue, it is incumbent on designers and developers to assess the risks associated with the operation of their systems and to specify the operator competencies needed to mitigate those risks. For more on the question of designer incompetence or negligence, see the discussion of "software malpractice" in Kroll (2018).

[68] The ISO standard on e-discovery, ISO/IEC 27050-3, does recognize the importance of expertise in applying advanced technologies in a search for documents responsive to a legal inquiry; see ISO/IEC 27050-3: *Information technology — Security techniques — Electronic discovery — Part 3: Code of practice for electronic discovery*, Geneva (2017), pp. 19-20.

[69] See, for example, ABA Model Rule 1, comment 8: "To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology, engage in continuing study and education and comply with all continuing legal education requirements to which the lawyer is subject." Available: https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_1_1_competence/comment_on_rule_1_1/. See also, The State Bar of California Standing Committee on Professional Responsibility and Conduct, Formal Opinion No. 2015-193. Available:

Law

https://www.calbar.ca.gov/Portals/0/documents/ethics/Opinions/CAL%202015-193%20%5B11-0004%5D%20(06-30-15)%20-%20FINAL.pdf.

[70] In the deliberations of the Law Committee of the 2018 Global Governance of AI Roundtable, the question of the competencies needed "in order to effectively operate and measure the efficacy of AI systems in legal functions that affect the rights and liberty of citizens" was cited as one of the considerations that "appear to be most overlooked in the current public dialogue." See "Global Governance of AI Roundtable: Summary Report 2018," World Government Summit, 2018: p. 7. Available: https://www.worldgovernmentsummit.org/api/publications/document?id=ff6c88c5-e97c-6578-b2f8-ff0000a7ddb6.

[71] See A. G. Ferguson, "Policing Predictive Policing," Washington University Law Review, vol. 94, no. 5, 2017: 1109, 1172. Available: https://openscholarship.wustl.edu/law_lawreview/vol94/iss5/5/.

[72] In addition, a lack of competence in interpreting the results of a statistical exercise can (and often does) result in an incorrect conclusion (on the part of a party to a dispute or of a judge seeking to resolve a dispute). For example, in *In re: Biomet*, a judge addressing a discovery dispute interpreted the statistical data provided by the producing party as indicating that the producing party's retrieval process had left behind "a comparatively modest number" of responsive documents, when the statistical evidence showed, in fact, that a substantial number of responsive documents had been left behind. See *In re: Biomet M2a Magnum Hip Implant Prods. Liab. Litig.*No. 3:12-MD-2391 (N.D. Ind. April 18, 2013).

[73] For example, a prior violent conviction may be weighted equally, whether the violent act was a shove or a knife attack. See Human Rights Watch. "Q & A: Profile Based Risk Assessment for US Pretrial Incarceration, Release Decisions," June 1, 2018. Available: https://www.hrw.org/news/2018/06/01/q-profile-based-risk-assessment-us-pretrial-incarceration-release-decisions.

[74] Bias can be introduced in a number of ways: via the features taken into consideration by the algorithm, via the nature and composition of the training data, via the design of the validation protocol, and so on. A competent operator will be alert to and assess such potential sources of bias.

[75] Among the conditions may be, for example, that the results of the system are to be used only to provide guidance to the human decision maker (e.g., judge) and should not be taken as, in themselves, dispositive.

[76] Given that the effective functioning of a legal system is a matter of interest to the whole of society, it is important that all members of a society be provided with access to the resources needed to understand when and how A/IS are applied in support of the functioning of a legal system.

[77] Among the topics covered by such training should be the potential for "automation bias" and ways to mitigate it. See L. J. Skitka, K. Mosier, and M. D. Burdick, "Does automation

bias decision-making?" International Journal of Human-Computer Studies, vol. 51, no. 5, pp. 991-1006, 1999. Available: https://doi.org/10.1006/ijhc.1999.0252; L. J. Skitka, K. Mosier, and M. D. Burdick, "Accountability and automation bias," International Journal of Human-Computer Studies, vol. 52, no. 4, pp. 701-717, 2000. Available: https://doi.org/10.1006/ijhc.1999.0349.

[78] Some government agencies are working toward creating a more effective partnership between the skills found in technology start-ups and the skills required of legal practitioners. See Legal Innovation Zone. "Ryerson's Legal Innovation Zone Announces Winners of AI Legal Challenge," March 26, 2018. Available: http://www.legalinnovationzone.ca/press_release/ryersons-legal-innovation-zone-announces-winners-of-ai-legal-challenge/.

[79] See Amazon. "Amazon Rekognition." https://aws.amazon.com/rekognition/ (2018).

[80] See E. Dwoskin, "Amazon is selling facial recognition to law enforcement—for a fistful of dollars." *Washington Post*, May 22, 2018. Available: https://www.washingtonpost.com/news/the-switch/wp/2018/05/22/amazon-is-selling-facial-recognition-to-law-enforcement-for-a-fistful-of-dollars/?noredirect=on&utm_term=.07d9ca13ab77.

[81] See, for example, J. Stanley, "FBI and Industry Failing to Provide Needed Protections for Face Recognition." *ACLU—Free Future*, June 15, 2016. Available: https://www.aclu.org/blog/privacy-technology/surveillance-technologies/fbi-and-industry-failing-provide-needed.

[82] It is also the case that, among the false positives, nonwhite members of Congress were overrepresented relative to their proportion in Congress as a whole, perhaps indicating that the accuracy of the technology is, to some degree, race-dependent. Without knowing more about the composition of the mugshot database, however, we cannot assess the significance of this result.

[83] See J. Snow, "Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots." *ACLU—Free Future*, July 26, 2018. Available: https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28. See also R. Brandom, "Amazon's facial recognition matched 28 members of Congress to criminal mugshots." *The Verge*, July 26, 2018. Available: https://www.theverge.com/2018/7/26/17615634/amazon-rekognition-aclu-mug-shot-congress-facial-recognition.

[84] See "Amazon Rekognition Developer Guide." Amazon, p. 131, 2018. Available: https://docs.aws.amazon.com/rekognition/latest/dg/rekognition-dg.pdf. Also see K. Tenbarge, "Amazon Responds to ACLU's Highly Critical Report of Rekognition Tech," Inverse, July 26, 2018. Available: https://www.yahoo.com/news/amazon-responds-aclu-apos-highly-160000264.html.

[85] The story also highlights the question of accountability, illustrating how the principles discussed in this report intersect with and complement each other.

Law

[86] Of course, competent use does not preclude use for bad ends (e.g., government surveillance that impinges on human rights). The principle of competence is one principle in a set that, collectively, is designed to ensure the ethical application of A/IS. See the EAD chapter "General Principles".

[87] Developing "well grounded" guidelines will typically require that the creators of A/IS gather input from both those operating the technology and those affected by the technology's operation.

[88] The use of facial recognition technologies by security and law enforcement agencies raises issues that extend beyond the question of operator competence. For further discussion of such issues, see C. Garvie, A. M. Bedoya, and J. Frankle, "The Perpetual Line-Up: Unregulated Police Face Recognition in America," *Georgetown Law, Center on Privacy & Technology*, October 18, 2016, Available: https://www.perpetuallineup.org/.

[89] As noted above, some professional organizations, such as the ABA, have begun to recognize in their codes of ethics the importance of technological competence, although the guidance does not yet address A/IS specifically.

[90] Including those engaged in the procurement and deployment of a system means that those acquiring and authorizing the use of a system can share in the responsibility for its results. For example, in the case of A/IS deployed in the service of the courts, this could be the judiciary; in the case of A/IS deployed in the service of law enforcement, this could be the agency responsible for the enforcement of the law and the administration of justice; in the case of A/IS used by a party to legal proceedings, this could be the party's counsel.

[91] J. New and D. Castro, "How Policymakers Can Foster Algorithmic Accountability."*Information Technology & Innovation Foundation,* p. 5, 2018. Available: https://www.itif.org/publications/2018/05/21/how-policymakers-can-foster-algorithmic-accountability.

[92] Included among possible "causes" for an effect are not only the decision-making pathways of algorithms but also, importantly, the decisions made by humans involved in the design, development, procurement, deployment, operation, and validation of effectiveness of A/IS.

[93] The challenge, moreover, is one not only of assigning responsibility, but of assigning levels of responsibility (a task that could benefit from a neutral model that could consider how much interaction and influence each stakeholder has in every decision).

[94] Scherer (2016): 372. In addition to diffuseness, Scherer identifies discreetness, discreteness, and opacity as features of the design and development of A/IS that make apportioning responsibility for their outcomes a challenge for regulators and courts.

[95] In answering these questions, it will be important to keep in mind the distinction between responsibility (a factual question) and ultimate accountability (a normative question). In the case of the example under discussion, there may be multiple individuals who have

Law

some practical responsibility for the sentence given, but the normative framework may place ultimate accountability on the judge. Before normative accountability can be assigned, however, pragmatic responsibilities must be clarified and understood. Hence the focus, in this section, on clarifying lines of responsibility so that ultimate accountability can be determined.

[96] If effectiveness is measured against statistics that themselves may represent human bias (e.g., arrest rates), then the effectiveness measures may just reflect and reinforce that bias.

[97] "'The algorithm did it' is not an acceptable excuse if algorithmic systems make mistakes or have undesired consequences, including from machine-learning processes." See "Principles for Accountable Algorithms and a Social Impact Statement for Algorithms." FAT/ML Resources. www.fatml.org/resources/principles-for-accountable-algorithms.

[98] See Langewiesche, W. 1998. "The Lessons of ValuJet 592". *Atlantic Monthly*. 281: 81-97; S. D. Sagan. *Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton University Press, 1995.

[99] For a discussion of the role of explanation in maintaining accountability for the results of A/IS and of the question of whether the standards for explanation should be different for A/IS than they are for humans, see F. Doshi-Velez, M. Kortz, R. Budish, C. Bavitz, S. J. Gershman, D. O'Brien, S. Shieber, J. Waldo, D. Weinberger, and A. Wood, Accountability of AI Under the Law: The Role of Explanation (November 3, 2017). Berkman Center Research Publication Forthcoming; Harvard Public Law Working Paper No. 18-07. Available: https://ssrn.com/abstract=3064761 or http://dx.doi.org/10.2139/ssrn.3064761.

[100] Also, gaining access to that information should not be unduly burdensome.

[101] Those developing a model for accountability for A/IS may find helpful guidance in considering models of accountability used in other domains (e.g., data protection).

[102] For a discussion of how such policies might be implemented in accordance with protocols for information governance, see J. R. Baron and K. E. Armstrong, "The Algorithm in the C-Suite: Applying Lessons Learned and Information Governance Best Practices to Achieve Greater Post-GDPR Algorithmic Accountability," in *The GDPR Challenge: Privacy, Technology, and Compliance In An Age of Accelerating Change*, A. Taal, Ed. Boca Raton, FL: CRC Press, forthcoming.

[103] These inquiries can be supported by technological tools that may provide information essential to answering questions of accountability but that do not require full transparency into underlying computer code and may avoid the necessity of an intrusive audit; see Kroll et al. (2017). Among the tools identified by Kroll and his colleagues are: software verification, cryptographic commitments, zero-knowledge proofs, and fair random choices. While the use of such tools may avoid the limitations of solutions such as transparency and audit, they do require that creators of A/IS design their systems so that they will be compatible with the application of such tests.

Law

[104] Certifications may include, for example, professional certifications of competence, but also certifications of compliance of processes with standards. An example of a certification program specifically addressing A/IS is *The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)*, https://standards.ieee.org/industry-connections/ecpais.html.

[105] This means that A/IS used in legal systems will have to be defensible in courts. The margin of error will have to be low or the use of A/IS will not be permitted.

[106] It is also the case that evidence produced by A/IS will be subject to chain-of-custody rules, as are other types of forensic evidence, to ensure integrity, confidentiality, and authenticity.

[107] See for instance Art. 22(1) Regulation (EU) 2016/679.

[108] Human dignity, as a core value protected by the United Nations Universal Declaration of Human Rights, requires us to fully respect the personality of each human being and prohibits their objectification.

[109] This concern is reflected in Principle 5 of the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment, recently published by the Council of Europe's European Commission for the Efficiency of Justice (CEPEJ). Principle 5 ("Principle 'Under User Control': preclude a prescriptive approach and ensure that users are informed actors and in control of the choices made") states, with regard to professionals in the justice system that they should "at any moment, be able to review judicial decisions and the data used to produce a result and continue not to be necessarily bound by it in the light of the specific features of that particular case," and, with regard to decision subjects, that he or she must "be clearly informed of any prior processing of a case by artificial intelligence before or during a judicial process and have the right to object, so that his/her case can be heard directly by a court." See CEPEJ, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment* (Strasbourg, 2018), p. 10.

[110] J. Tashea, Calculating Crime: Attorneys are Challenging the Use of Algorithms to Help Determine Bail, Sentencing and Parole, ABA Journal (March 2017).

[111] Loomis v. Wisconsin, 68 WI. (2016).

[112] *Id.* at pp. 46-66.

[113] R. Wexler, Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System, Stanford Law Review, 2018.

[114] Malenchik v. State, 928 N.E.2d 564, 574 (Ind. 2010).

[115] People *v.* Chubbs CA2/4, B258569 (Cal. Ct. App. 2015).

[116] U.S. v. Ocasio, No. 3:11-cr-02728-KC, slip op. at 1-2, 11-12 (W.D. Tex. May 28, 2013).

[117] U.S. v. Johnson, No. 1:15-cr-00565-VEC, order (S.D.N.Y., June 7, 2016).

[118] Indeed, without transparency, there may, in some circumstances, be no means for even knowing whether an error that needs to be corrected was committed. In the case of A/IS

Law

applied in a legal system, an "error" can mean real harm to the dignity, liberty, and life of an individual.

[119] *Fairness* (as well as *bias*) can be defined in more than one way. For purposes of this discussion, a commitment is not made to any one definition—and indeed, it may not be either desirable or feasible to arrive at a single definition that would be applied in all circumstances. For purposes of this discussion, the key point is that transparency will be essential in building informed trust in the fairness of a system, regardless of the specific definition of *fairness* that is operative.

[120] To the extent permitted by the normal operation of the A/IS: allowing for, for example, variation in the human inputs to a system that may not be eliminated in any attempt at replication.

[121] With regard to information explaining how a system arrived at a given output, GDPR makes provision for a decision subject's right to an explanation of algorithmic decisions affecting him or her: automated processing of personal data "should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision." GDPR, Recital 71.

[122] Even among sensitive data, some data may be more sensitive than others. See I. Ajunwa, "Genetic Testing Meets Big Data: Tort and Contract Law Issues," 75 *Ohio St. L. J.* 1225 (2014). Available: https://ssrn.com/abstract=2460891.

[123] See A. Baker, "Updated N.Y.P.D. Anti-Crime System to Ask: 'How We Doing?'" *New York Times*, May 8, 2017, https://www.nytimes.com/2017/05/08/nyregion/nypd-compstat-crime-mapping.html; S. Weichselbaum, "How a 'Sentiment Meter' Helps Cops Understand Their Precincts," *Wired*, July 16, 2018. Available: https://www.wired.com/story/elucd-sentiment-meter-helps-cops-understand-precincts/.

[124] This table is a preliminary draft and is meant only to illustrate a useful tool for facilitating reasoning about who should have access to what information. Other categories of stakeholder and other categories of information (e.g., the identity and nature of the designer/manufacturer of the A/IS, the identity and nature of the investors backing a particular system or company) could be added as needed.

[125] For discussions of these two dimensions of explanation, see S. Wachter, et al. (2017). "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation"; A. Selbst, and S. Barocas, The Intuitive Appeal of Explainable Machines.

[126] Wexler, Rebecca. 2018. "Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System". *Stanford Law Review*. 70 (5): 1342-1429; Tashea, Jason. "Federal judge releases DNA software source code that was used by New York City's crime lab." *ABA Journal* (2017). http://www.abajournal.com/news/article/federal_judge_releases_dna_software_source_code.

[127] Or, if two approaches are found to be, for practical purposes, equally effective, the simpler, more easily explained approach may be preferred.

Law

[128] For a discussion of the limits of transparency and of alternative modes of gaining actionable answers to questions of verification and accountability, see J.A. Kroll, J. Huey, S. Barocas, E.W. Felten, J.R. Reidenberg, D.G. Robinson, H. Yu, "Accountable Algorithms" (March 2, 2016). *University of Pennsylvania Law Review*, Vol. 165, 2017 Forthcoming; Fordham Law Legal Studies Research Paper No. 2765268. Available at SSRN: https://ssrn.com/abstract=2765268. See also J.A. Kroll, The fallacy of inscrutability, *Phil. Trans. R. Soc.* A 376: 20180084. http://dx.doi.org/10.1098/rsta.2018.0084 (Note p. 9: "While transparency is often taken to mean the disclosure of source code or data, possibly to a trusted entity such as a regulator, this is neither necessary nor sufficient for improving understanding of a system, and it does not capture the full meaning of transparency.")

[129] In particular with respect to due process, the current dialogue on the use of A/IS centers on the tension between the need for transparency and the need for the protection of intellectual property rights. Adhering to the principle of Effectiveness as articulated in this work can substantially help in defusing this tension. Reliable empirical evidence of the effectiveness of A/IS in meeting specific real-world objectives may foster informed trust in such A/IS, without disclosure of proprietary or trade secret information.

[130] S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR," SSRN Electronic Journal, p. 5, 2017 for the example cited.

[131] W. L. Perry, B. McInnis, C. C. Price, S. C. Smith, and J. S. Hollywood, "Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations," The RAND Corporation, pp. 67-69, 2013.

[132] Support from the University of Memphis was led by Richard Janikowski, founding Director of the Center for Community Criminology and Research (School of Urban Affairs and Public Policy, the University of Memphis) and the Shared Urban Data System (The University of Memphis).

[133] E. Figg, "The Legacy of Blue CRUSH," High Ground, March 19, 2014.

[134] Figg, "Legacy."

[135] Nucleus Research, *ROI Case Study: IBM SPSS—Memphis Police Department*, Boston, Mass., Document K31, June 2010. Perry et al., *Predictive Policing*, 69.

[136] Figg, "Legacy."

[137] Figg, "Legacy."

[138] See: AI Now, *Algorithmic Accountability Policy Toolkit*, p. 12, Oct. 2018. Available: https://ainowinstitute.org/aap-toolkit.pdf; D. Robinson and L. Koepke, *Stuck in a Pattern: Early evidence on "predictive policing" and civil rights*, Upturn, Aug. 2016. Available: https://www.upturn.org/reports/2016/stuck-in-a-pattern/; S. Brayne, "Big Data Surveillance: The Case of Policing," American Sociological Review, 2016. Available: https://journals.sagepub.com/doi/10.1177/0003122417725865; A. G. Ferguson, "Policing Predictive Policing,"

Law

Washington University Law Review, vol. 94, no. 5, 2017. Available: https://openscholarship.wustl.edu/law_lawreview/vol94/iss5/5/; K. Lum and W. Isaac, "To predict and serve?" *Significance* 2016. Available: https://rss.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1740-9713.2016.00960.x; B. J. Jefferson, "Predictable Policing: Predictive Crime Mapping and Geographies of Policing and Race," Annals of the American Association of Geographers, vol. 108, no. 1, pp. 1-16, 2018. Available: https://doi.org/10.1080/24694452.2017.1293500.

[139] For a discussion of the criteria that may define a "high-crime area," and so potentially license more intrusive policing, see A. G. Ferguson and D. Bernache, "The 'High-Crime Area' Question: Requiring Verifiable and Quantifiable Evidence for Fourth Amendment Reasonable Suspicion Analysis," American University Law Review, vol. 57, pp. 1587-1644.

[140] While A/IS, if misapplied, may perpetuate bias, it holds at least the potential, if applied with appropriate controls, to reduce bias. For a study of how an impersonal technology such as a red light camera may reduce bias, see R. J. Eger, C. K. Fortner, and C. P. Slade, "The Policy of Enforcement: Red Light Cameras and Racial Profiling," *Police Quarterly*, pp. 1-17, 2015. Available: http://hdl.handle.net/10945/46909.

[141] See, for example: J. Tashea, "Estonia considering new legal status for artificial intelligence," ABA Journal, Oct. 20, 2017, and European Parliament Resolution of Feb. 16, 2017.

[142] See Legal Entity, Person, *in* B. Bryan A. Garner, Black's Law Dictionary, 10th Edition. Thomas West, 2014.

[143] J. S. Nelson, "Paper Dragon Thieves." Georgetown Law Journal 105 (2017): 871-941.

[144] M. U. Scherer, "Of Wild Beasts and Digital Analogues: The Legal Status of Autonomous Systems." Nevada Law Journal 19, forthcoming 2018.

[145] See M. U. Scherer, "Of Wild Beasts and Digital Analogues: The Legal Status of Autonomous Systems." Nevada Law Journal 19, forthcoming 2018; J. F. Weaver. Robots Are People Too: How Siri, Google Car, and Artificial Intelligence Will Force Us to Change Our Laws. Santa Barbara, CA: Praeger, 2013; L. B. Solum. "Legal Personhood for Artificial Intelligences." North Carolina Law Review 70, no. 4 (1992): 1231–1287.