DAKSH

ALGORITHMIC **ACCOUNTABILITY** IN THE **JUDICIARY**

Authors

Amulya Ashwathappa

Leah Verghese

Sandhya P.R

Siddharth Mandrekar Rao

Acknowledgements

The authors would like to thank Surya Prakash B.S. and Chockalingam Muthian for their valuable suggestions to this paper. The authors would also like to thank Satyajit Amin, who was an intern with DAKSH for his research assistance.

For any queries and clarifications regarding this paper please email info@dakshindia.org

To follow more of DAKSH's work, please visit www.dakshindia.org

You can also follow us on our social media handles:

@dakshimpact

@daksh_india

DAKSH

Contents

INTRODUCTION 6

SECTION 2:

Terms, concepts, and definitions **9**

- Artificial Intelligence (AI) 9
- Algorithm 10
- Machine Learning 11

SECTION 3:

Functions and capabilities of advanced algorithmic tools and systems 13

- Predictive tools 13
- Case law research 13
- Risk assessment tools 13
- Artificial intelligence, algorithms and dispute resolution 13
- Language recognition 14
- Digital file management 14
- Vision (perception) 14

SECTION 4:

Ethical principles **16**

- Transparency 16
- Bias 18
- Security 24
- Accountability 24
- Accessibility and Inclusion 25
- Privacy 25
- Right to a human decision 28

SECTION 5:

International Experience **31**

United Kingdom 31

European Union (eu) 31

China 33

Netherlands 34

France 34

Canada 35

SECTION 6:

An institutional arrangement for the regulation of algorithms **38**

The necessity of a regulatory body under the judiciary 38

Regulatory and organisational challenges resulting from the use of algorithms in the judiciary 41

Defining the regulatory unit 41

Rights and responsibilities 42

Composition - competencies, representation 43

Regulatory activities 45

Standard-setting 46

Selection, certification, and audit of algorithmic systems 46

Grievance Redressal, Ownership, Liability 48

Public Consultation and Engagement 49

Training and education within the judiciary 49

Implementation 49

Introduction

The Indian judiciary has undertaken sustained efforts to modernise and digitise the systems for managing records and conducting various processes that are part of the life cycle of a case.[1] Among the most recent developments is the introduction of applications that members of the judiciary, including the recently retired Chief Justice of India (CJI), Justice S.A. Bobde, have referred to as 'Artificial Intelligence' (AI). At present, these include the 'Supreme Court Vidhik Anuvaad Software' (SUVAS),[2] which is used to translate judgments of courts from English into Indian languages; and the 'Supreme Court Portal for Assistance in Courts Efficiency' (SUPACE), a tool to help judges conduct legal research.[3]

While the functions of these particular applications are clear, the Supreme Court has not published any details on their internal workings or whether any rules, guidelines, or policies have been instituted to regulate the development and use of such applications.[4] The information available is primarily from press releases and judges' statements to the effect that AI will not substitute the decision-making capacity of judges.[5] This raises multiple questions. Administrative decisions have an impact on judicial decision-making, what steps can be and have been taken to act upon this guarantee? How is the judiciary drawing boundaries between AI and related concepts such as automation and machine learning? What measures are being taken to ensure accountability for adverse events resulting from deploying these AI technologies? Perhaps most importantly, what does the judiciary mean when they refer to AI, and by what means can it be regulated?

This paper proposes that many of these questions can be adequately addressed by regulating algorithms, for two reasons. One is that algorithms form the building blocks of digital technology and are the mechanisms underpinning any technology that can be considered to be AI, irrespective of whether they can be regarded as 'intelligent'. The other reason is that the use of the term 'algorithm' is backed by a broad consensus, and does not raise contentious definitional debates, and its usage has been more consistent.[6] Algorithms form the building blocks of all automated processes, They serve as a better starting point to explore the use of automation in the administration of justice, as they direct

1 E-Committee, Supreme Court of India. 'eCourts Project Phase II Objectives Accomplishment Report As per Policy Action Plan Document'. E-Courts, available online at https://ecourts.gov.in/ecourts_home/static/manuals/Objectives%20Accomplishment%20Report-eCourts-final_copy.pdf (accessed on 4 May, 2021)

2 https://main.sci.gov.in/pdf/Press/press%20release%20for%20law%20day%20celebratoin.pdfs

3 https://www.livemint.com/news/india/cji-s-a-bobde-welcomes-ai-system-to-assist-judges-in-legal-research-11617725127705.html

4 As of May 2021

5 https://www.indiatoday.in/india/story/supreme-court-india-sc-ai-artificial-intellegence-portal-supace-launch-1788098-2021-04-07

6 Donald E. Knuth. 'The Art of Computer Programming, Vol. I/Fundamental Algorithms'. For the definition of algorithm used in this paper, see the next section.

every use of technology in place of a task that was previously (or is currently) performed by humans. As a result of this choice, we refer primarily to algorithms throughout this paper, rather than 'AI', unless we refer to sources that use the term 'AI'.

The digitisation of judicial processes presents a unique set of opportunities and challenges. While doing so opens up possibilities to increase access to justice, there are numerous unaddressed concerns, particularly with regard to transparency and discriminatory patterns that can result, in addition to other ethical issues. This paper proposes an approach to creating a regulatory framework for algorithmic accountability that will address these challenges, emphasising the issues raised by the deployment of advanced algorithms that are substituted for more complex acts of human decision-making.

Section 2 provides a discussion of some basic concepts, clarifies our use of them, and indicates how they would be applied to a judicial context. Section 3 describes the functions and capabilities of algorithms in this context. Section 4 discusses the ethical concerns regarding algorithmic accountability in the use of advanced algorithmic decision-making systems in the judicial context. Section 5 provides an overview of the efforts of other jurisdictions to incorporate and regulate algorithms in a judicial context. Finally, Section 6 describes the regulatory challenges for their use in this context and discusses how regulation can address them.

Section 2:

Terms, concepts, and definitions

This chapter defines terms and concepts relevant to algorithmic accountability, and where necessary, clarifies our use of them in this paper. We are doing so for two reasons. Firstly, the ambiguity about many of these terms makes it necessary to clarify our interpretation of them. Secondly, our choice of a given interpretation has implications for the nature of technologies and use cases that this paper proposes. This would affect the proposed policy and regulation, given the ethical and legal implications of algorithmic technologies in the judicial sector.

Artificial Intelligence (AI)

Attempts to define AI have been made since the inception of the field, but with little consensus. Defining AI requires consensus on the meaning of intelligence, which is highly contested subject. How researchers define AI has historically determined the path of research they pursue and the technological capabilities that ultimately result from their work. We provide an overview of these definitions and how they influence approaches in AI development. This shows the wide range of interpretations of the term AI and how these are problematic in the judicial context. We group together technologies conventionally regarded as AI, irrespective of whether they meet philosophical and psychological criteria for intelligence.

Russell and Norvig[7] recommend defining AI as the ability to act rationally rather than in terms of achieving 'thought' or in terms of human behaviour. They argue that it is still possible to act rationally when information is uncertain and because rationality can be logically and mathematically defined, making this approach easier to apply to computers.

In addition to the general challenges of defining AI, there are additional challenges to defining it to prescribe, mandate, and regulate its use in law and the judicial system. As later sections will reveal, there is a diverse range of technologies, capabilities, and developmental approaches that fall under the umbrella of AI. This raises the question of whether the definition, or even the term itself, is an adequate grouping of technologies for policy and regulatory purposes. For judicial applications of AI we need a precise definition. Others have attempted to resolve this dilemma by defining the aspects of AI that are seen to need regulation. This avoids reliance on contested abstractions such as 'intelligence'.

The Centre for Internet and Society defines AI as 'a dynamic learning system that can be used in decision making and action'.[8] This captures its learning and decision-making capabilities, an advantage over the earlier definitions since these recent developments are the main cause of concern from a regulatory perspective. However, it omits mention of the lack of a human in this system, or the ability to act, learn, and make decisions without human intervention. Jacob Turner defines it as 'the ability of a non-natural entity to make choices by an evaluative process'.[9] Though this definition does not mention learning capabilities, its broader scope (inclusive of non-learning algorithms) and mention of non-natural entities both emphasises a degree of removal from human beings, while allowing for the future possibility of AI creating other AI.

While Russell and Norvig advocate the 'rational agents' approach to defining AI, Jonas

7 Stuart J. Russell and Peter Norvig. 2010. Artificial Intelligence: A Modern Approach (Third Edition). Essex, UK: Pearson Education Ltd.

8 Geethanjali Jujjavarapu, Elonnai Hickok, Amber Sinha, Shweta Mohandas, Sidharth Ray, Pranav M. Bidare, and Mayank Jain. 2018. 'AI and the Manufacturing and Services Industry in India.' 6 January. The Center for Internet and Society, India. URL: https://cisindia. org/internetgovernance/files/AIManufacturingandS ervices_Report_02. pdf.

9 Jacob Turner. 2018. Robot rules: Regulating artificial intelligence. Cham, Switzerland: Springer., Matthew U. Scherer. 2015. 'Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies.' Harv. JL & Tech. 29: 353.

Schuett observes that replacing "intelligence" with "goal" simply substitutes one term that is hard to define for another, since the term "goal" implies a degree of self-awareness.[10] When approaching this problem from a regulator's perspective, Schuett describes the attributes that are necessary for a satisfactory legal definition.[11] These are the appropriate level of inclusiveness in relation to the regulatory goal, precision, how easy it is for entities to understand whether their behaviour is compliant, how easy it is to assess whether a case meets the definition and permanence. He claims that some definitions are over-inclusive, such as Russell and Norvig's, and others can be under-inclusive, such as Turing's, and rejects them in legal applications on these grounds. Instead, Schuett proposes avoiding defining AI itself and recommends that regulation instead be adapted to specific technologies, how they work, the use cases they are employed in, the risks they pose, and the properties of these technologies that create these risks.[12]

It is clear that there is a lack of consensus regarding general definitions and how to define AI for regulatory purposes.

Algorithm

Algorithms are a process or procedure and the set of steps to be followed in solving a mathematical or logical problem. They are used to tell computers what to do and are used in every task computers perform, from the simplest to the most complex.[13] Knuth describes algorithms as "a finite set of rules which gives a sequence of operations for solving a specific type of problem..", which possesses the following characteristics:[14]

1. **Finiteness** –it terminates after a finite number of steps
2. **Definiteness** – each step is precisely defined
3. **Inputs** – consisting of a specified set numbering zero or more
4. **Outputs** – consisting of specified set numbering one or more, and having a definite relation to the inputs
5. **Effectiveness** – meaning that all operations can be performed "exactly and in a finite length of time"[15]

Machine Learning

Machine learning algorithms are a class of algorithms that can "learn" to perform better on a given task through experience.[16] Mitchell provides the following definition:

> "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."[17]

Many definitions of machine learning assert that being able to perform a task without being explicitly told how to do so is its defining characteristic.[18] Machine learning algorithms are "algorithms that make other algorithms"[19] and are especially useful for solving problems for which we do not have an algorithm but for which **data** pertaining to the problem is available. Such data provides "examples of some phenomenon" that machine learning

10 Stuart J. Russell and Peter Norvig. 2010. Artificial Intelligence: A Modern Approach (Third Edition). Essex, UK: Pearson Educ

11 Jonas Schuett. 2019. 'A Legal Definition of AI.' Available at SSRN 3453632

12 Jonas Schuett. 2019. 'A Legal Definition of AI.' Available at SSRN 3453632

13 Pedro Domingos. 2015. The master algorithm: How the quest for the ultimate learning machine will remake our world. Basic Books .

14 Donald E. Knuth. 1973. The Art of Computer Programming, Vol. I/ Fundamental Algorithms. Reading, USA: Addison-Wesley Publishing Company.

15 Donald E. Knuth. 'The Art of Computer Programming, Vol. I/Fundamental Algorithms'.

16 Peter Flach. Machine learning: the art and science of algorithms that make sense of data. Cambridge University Press, 2012,

17 Tom M. Mitchell. 1997. Machine Learning. Burr Ridge, USA: McGraw Hill.

18 Domingos, Pedro. The master algorithm: How the quest for the ultimate learning machine will remake our world.

19 Domingos, Pedro. The master algorithm: How the quest for the ultimate learning machine will remake our world.

algorithms use to program algorithms that perform a particular task well.[20]

These algorithms are built upon "models" of relationships between the outcome sought and the factors which influence it, whose interrelationship is governed by "parameters" that can be modified as the algorithm is refined through experience.[21] The modifications that the machine learning algorithm makes to these parameters are those that optimise a "performance criterion", which indicates how well the algorithm is performing the task assigned to it.

A significant factor for the recent success of machine learning is that the digitisation of institutions and processes in the public and private sectors have generated large volumes of data for machine learning algorithms to use. The E-Courts Mission Mode Project has similarly generated a large volume of data on court cases in district courts. Machine learning algorithms could use this data to perform tasks such as identifying cases that are likely to take longer to get disposed. The advantage that machine learning has over other conventional statistical analyses is that it does not require statistical assumptions about the data.[22] This enables the detection of patterns that would be very difficult for humans to detect, as they may be arbitrary or require processing large volumes of data. They also automatically improve over time, whereas using conventional statistical methods to achieve similar insights may require repeated human intervention. While this has yielded great advances in capabilities, the fact that these algorithms learn on their own makes it hard to understand the reasoning behind their conclusions, which is problematic in domains where their use has ethical implications.[23] These algorithms are therefore sometimes referred to as 'black boxes'. This is especially the case for the judiciary, where the use of algorithms, even for tasks such as the listing of cases, would raise questions of fairness if these algorithms were opaque.[24]

20 Andriy Burkov. 2019. The hundred-page machine learning book. Available at http://themlbook.com/

21 Alpaydin, Ethem. Introduction to machine learning. MIT press, 2020.

22 For example, assumptions aobut underlying distributions within the data, particularly with regard to functional form, like linearity

23 Amina Adadi and Mohammed Berrada. 2018. 'Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI).' IEEE Access 6 : 52138-52160.

24 For more on 'explainable AI', see Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. 2017. 'Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models.' arXiv preprint, arXiv:1708.08296 Amina Adadi and Mohammed Berrada. 2018. 'Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI).' IEEE Access 6 : 52138-52160., Tutt, Andrew. "An FDA for algorithms." Admin. L. Rev. 69 (2017): 83., Bernhard Waltl and Roland Vogl. 2018. 'Explainable artificial intelligence the new frontier in legal informatics.' Jusletter IT 4 : 1-10.Deeks, Ashley. "The Judicial Demand for Explainable Artificial Intelligence." Columbia Law Review 119, no. 7 (2019): 1829-1850.

Section 3:

Functions and capabilities of advanced algorithmic tools and systems

Algorithmic tools have been implemented in various jurisdictions to assist stakeholders in performing their tasks effectively. Below are examples of some of the artificial intelligence tools that have been developed in the law and justice space.

Predictive tools

Many companies have taken advantage of big data to analyse the profiles of judges and to predict the outcomes of cases.[25] Such tools can be used to predict chances of winning a case, estimate the cost of litigation, and provide other analyses to litigants and lawyers. Predictive tools can provide insights by linking multiple variables with the available data to determine patterns. Scholars have warned that the use of big data in predictive analysis can lead to oversimplification of legal problems. Predictive analysis also limits the scope of use to past actions and cannot predict novel possibilities.[26]

Case law research

Artificial intelligence tools can process not just written documents, but also audio and video files.[27] Some case research tools allow lawyers to transcend text search results. For example, these tools allow lawyers to deep dive into a particular case type, see decisions that were published in a specific time and categorise them.[28]

25 Daniel Faggella. 2020. 'AI in Law and Legal Practice – A Comprehensive View of 35 Current Applications', Emerge Artificial Intelligence Research, 14 March, available online at https://emerj.com/ai-sector-overviews/ai-in-law-legal-practice-current-applications/ (accessed on 28 September 2020).

26 Caryn Devins, Teppo Felin, Stuart Kauffman and Roger Koppl 'The Law and Big Data', Cornell Journal of Law and Public Policy. 27(357), available online at https://www.lawschool.cornell.edu/research/JLPP/upload/Devins-et-al-final.pdf (accessed on 28 September 2020).

27 Judge Herbert B. Dixon Jr. 2020. 'What Judges and Lawyers Should Understand About Artificial Intelligence Technology' American Bar Association, 3 February, available online at https://www.americanbar.org/groups/judicial/publications/judges_journal/2020/winter/what-judges-and-lawyers-should-understand-about-artificial-intelligence-technology/ (accessed on 28 September 2020).

28 Faggella, 'AI in Law and Legal Practice – A Comprehensive View of 35 Current Applications'.

Risk assessment tools

Risk assessment tools are used in the criminal justice system to make decisions on who is eligible for parole, bail etc. The prisoner is assessed based on factors like race, age, sex, prior criminal history, socioeconomic conditions, psychological evaluation etc.[29] In the U.S., 'COMPAS', a risk assessment tool has been questioned for reinforcing the existing racial bias in their justice system.[30] Legal scholars have stressed on the fact that these tools should be 'fair and just' and should be trained to imbibe fundamental ethical principles.

Artificial intelligence, algorithms and dispute resolution

Many countries are using algorithms to resolve civil disputes, including traffic cases and to decide the quantum of fines. In the 'rule-based reasoning model,' the system is trained with the rules and the user provides the facts and the system will arrive at a decision.[31] In the case-based reasoning model the system is built on the system's experience, its ability to compare and articulate reasons from other outcomes. Case-based reasoning relies on the system's ability to use its previous successful decisions, and to recognise similar failures in advance, so they can be avoided in the future.[32] Around the world many countries have employed hybrids of these models to

29 Centre for Court Innovation. 2019.'Beyond the Algorithm: Pretrial Reform, Risk Assessment, and Racial Fairness', New York: Centre for Court Innovation, p. xx. Available online at https://www.courtinnovation.org/sites/default/files/media/documents/2019-06/beyond_the_algorithm.pdf (accessed on 28 September 2020).

30 Dixon Jr., 'What Judges and Lawyers Should Understand About Artificial Intelligence Technology';

Jeff Larson, Surya Mattu, Lauren Kirchner and Julia Angwin. 2016. 'How We Analyzed the COMPAS Recidivism Algorithm'. Pro Publica, 23 May, available online at https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (accessed on 28 September 2020).

31 Suzanne Van Arsdale. 2015. 'User Protections in Online Dispute Resolution' Harvard Negotiation Law Review 21(107).

32 Arno R. Lodder and John Zeleznikow. 2013. 'Artificial Intelligence and Online Dispute Resolution' in Mohamed Abdel Wahab, Ethan Katsh and Daniel Rainey (ed.), Online Dispute Resolution Theory and Practice, pp 73-94. Hague: Eleven International Publishing.

settle consumer, traffic, divorce cases, etc.[33]

Language recognition

Another useful application of artificial intelligence in courtrooms is the live transcription of hearings. Natural language processing is used to recognise what is said in courtrooms and record it. These tools are also able to process nuances in language.[34] The extent of the accuracy of the text depends on the quality of the voice and also the intuitiveness of the system.

Digital file management

Solutions for file management were initially based on simple software tools used to store and manage files. Big data, cloud storage and algorithms have changed how cases and documents are managed by court staff, judges, lawyers etc. The application of algorithms and artificial intelligence tools in this field has reduced the time spent on case files.[35] They can help in reducing the size of the files and can assist in retrieving files from unstructured data.[36] Optical Character Recognition (OCR) is a type of algorithm used by courts to manage document file management and is used during filing of cases, submissions and generally assisting litigants and lawyers.[37] 'Automated Docketing' is another form of artificial intelligence used to automatically identify the case and types of the cases and process them.[38]

33 Jeremy Barnett and Philip Treleaven. 2017. 'Algorithmic Dispute Resolution -The Automation of Professional Dispute Resolution Using AI and Blockchain Technologies' The Computer Journal, 61 (3): 299-408.

34 Joint Technology Committee. 2020. Introduction to AI for Courts. United States: Joint Technology Committee, p. X. Available online at https://www.ncsc.org/__data/assets/pdf_file/0013/20830/2020-04-02-intro-to-ai-for-courts_final.pdf (accessed on 28 September 2020).

35 Faggella, 'AI in Law and Legal Practice – A Comprehensive View of 35 Current Applications'.

36 Faggella, 'AI in Law and Legal Practice – A Comprehensive View of 35 Current Applications'.

37 Joint Technology Committee, Introduction to AI for Courts

38 Joint Technology Committee, Introduction to AI for Courts.

Vision (perception)

AI tools assisting in facial recognition are advanced tools that use stored images and videos to identify people based on the available data on the system.[39] In criminal investigations, video analytical tools can geotag people to a particular place and time. It can also create a 3D image that helps in recreating the crime scene.[40] Video analytics can be used during online hearings to analyse facial expressions, body posture, and gaze to assist the judiciary in the evaluation of litigants[41] and combat the fear of tutoring witness during online hearings.

39 IJIS Technology and Architecture Committee (ITAC). Artificial Intelligence in Justice and Public Safety.

40 Daniel Faggella. 2019. 'AI for Crime Prevention and Detection – 5 Current Applications', Emerge Artificial Intelligence Research, 2 February, available online at https://emerj.com/ai-sector-overviews/ai-crime-prevention-5-current-applications/ (accessed on 28 September 2020)

41 DAKSH. 2020. Video Conferencing in Indian Courts: A Pathway to the Justice Platform. Bengaluru: DAKSH.

Section 4:

Ethical principles

The advances in capabilities of algorithms in various fields have raised ethical questions regarding their use in various applications, and how to prevent them from harming people, whether inadvertently or maliciously. This section discusses the principles which are applicable to the regulation of algorithms, the ethical challenges that they pose, and their relevance to the use of algorithms in judicial information systems.

This section makes frequent reference to machine learning algorithms, which, as Section 2 indicated, are a sub-class of algorithms which are capable of improving their performance. Many of the ethical questions discussed here have emerged or grown in prominence specifically because of the characteristics of machine learning, such as their opacity, non-intuitive process of inference, and the difficulty of ascribing agency (and legal liability) to either an inanimate technology or human actors who use their inferences to make a decision impacting the lives and rights of others.

Transparency

This chain of logic as discussed above, commonly drives transparency concerns: observation produces insights that create the knowledge required to govern and hold systems accountable. Observation is understood as a diagnostic for ethical action, as observers with more access to the facts describing a system will better judge whether a system is working as intended and what changes are required. The more that is known about a system's inner workings, the more defensibly it can be governed and held accountable.[42] E.g. if an algorithm is scheduling cases for a particular judge, transparency would require the parameters the algorithm is using to perform its task and the model used to do so to be disclosed.

Primarily, transparency is a way to minimise harm and improve algorithms, though some sources underline its benefit for legal reasons or to foster trust. A few sources also link transparency to dialogue, participation and the principles of democracy. [43]

The leading cause of widespread and institutional distrust of the use of algorithms to perform tasks with serious consequences, particularly those that can do so with a degree of autonomy, is the lack of transparency resulting from the opacity of recent technologies such as artificial neural networks, and the means of producing their output. While the learning algorithm may be open and transparent, the model it produces may not be. This has implications for developing machine learning systems, but more importantly, for their safe deployment and accountability.[44] It may be necessary to access code to hold a system accountable, but seeing code is insufficient. Furthermore, system builders themselves are often unable to explain how a complex system works, which parts are essential for its operation, or how the ephemeral nature of computational representations are compatible with transparency laws. [45] In this context, the argument favouring retaining human actors is that they can be asked to explain their reasoning in reaching any given decision, which an algorithmic process cannot.

The difficulty of achieving explainability with the development of machine learning algorithms has led to the emergence of the

42 Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability, 974; Adrienne Yapo and Joseph Weiss. 2018. 'Ethical implications of bias in machine learning.' Proceedings of the 51st Hawaii International Conference on System Sciences. Available at https://doi.org/10.24251/HICSS.2018.668

43 Anna Jobin, Marcello Ienca and Effy Vayena, The global landscape of AI ethics guidelines

44 Internet Society. 2017. 'Artificial Intelligence and Machine Learning: Policy Paper.' Internet Society. Available online at https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper/

45 Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability, 981

concept of 'interpretability', a more loosely-defined term referring to the ability of domain experts to interpret the output of a machine learning model.[46] Part of the problem is that what is an acceptable level of interpretation varies based on the domain, the application, and the individual characteristics and beliefs of the interpreter. Some have attempted to propose approaches to defining interpretability based on characteristics, techniques, and elements of machine learning models and algorithms that are 'thought to confer interpretability'.[47]

The main argument against the strict enforcement of explainability of algorithms is the trade-off between explainability and accuracy.[48] Machine learning algorithms are capable of far greater performance than previous technologies,[49] but their output is much harder to predict. Many experts make the argument that accuracy, rather than transparency, is a much more important parameter in various fields (such as medicine).[50] However, this position will find little support in the legal context because the process is as important as the outcome in a legal proceeding. All parties to a dispute must be satisfied that the processes that lead to a judicial decision are fair, even administrative ones.[51] This may be extended to investigating officers and other agencies in criminal cases, where explanatory standards may be applied even to human officials.[52] The Supreme Court judgment in *Olga Tellis & Ors vs Bombay Municipal Corporation* quoted Laurence Tribe on this subject:

> *"Whatever its outcome, such a hearing represents a valued human interaction in which the affected person experiences at least the satisfaction of participating in the decision that vitally concerns her, and perhaps the separate satisfaction of receiving an explanation of why the decision is being made in a certain way. Both the right to be heard from, and the right to be told why, are analytically distinct from the right to secure a different outcome; these rights to inter change express the elementary idea that to be a person, rather than a thing is at least to be consulted about what is done with one."*
>
> ***Laurence H. Tribe, American Constitutional Law"*** [53]

Understanding the reasoning for a judgment is a critical element of procedural due process, and therefore any algorithms used in the judicial process must meet high standards of explainability. Using more transparent but less accurate algorithms is acceptable or avoiding them altogether until algorithms that meet standards of accuracy and transparency are both preferable to using an accurate but inscrutable algorithm in the judicial context. Frank Pasquale illustrates this trade-off and the importance of explainability:

46 Asking 'Why' in AI: Explainability of intelligent systems – perspectives and challenges

47 The Mythos of Model Interpretability Zachary C. Lipton

48 Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy

49 Domingos, Pedro. The master algorithm: How the quest for the ultimate learning machine will remake our world. Basic Books, 2015.

50 London, Alex John. "Artificial intelligence and black-box medical decisions: accuracy versus explainability." Hastings Center Report 49, no. 1 (2019): 15-21.

51 Winn, Peter A. "Online court records: Balancing judicial accountability and privacy in an age of electronic information." Wash. L. Rev. 79 (2004): 307. Conley, Amanda, Anupam Datta, Helen Nissenbaum, and Divya Sharma. "Sustaining privacy and open justice in the transition to online court records: A multidisciplinary inquiry." Md. L. Rev. 71 (2011): 772. 5. Morrison, Caren Myers. "Privacy, accountability, and the cooperating defendant: Towards a new role for internet access to court records." Vand. L. Rev. 62 (2009): 919

52 Brennan-Marquez, Kiel. "Plausible cause: Explanatory standards in the age of powerful machines." Vand. L. Rev. 70 (2017): 1249.

53 1986 AIR 180, 1985 SCR Supl. (2) 51

> *"A voice-parsing algorithm might predict Supreme Court votes much more cheaply than the justices and clerks arguing and writing out decisions. But no respectable legal system would substitute it for actual legal determinations, at whatever level of the justice system it might be deployed, because it cannot relate its rationale with reasons that have normative weight. Explainability matters because the process of reason-giving is intrinsic to juridical determinations—not simply one modular characteristic jettisoned as anachronistic once automated prediction is sufficiently advanced."* [54]

One of the most significant regulatory goals for the safe and responsible use of algorithms within the judiciary is to establish standards for the explainability of algorithms.[55] There should be a high level of consensus among all stakeholders regarding these standards and they should satisfy constitutional and jurisprudential principles, values, and doctrines.[56] Standards for explainability should also be flexible enough to account for the contexts in which algorithms are used and how people's demands for explainability may vary based on these contexts. The design of these standards should ensure that compliance with them should be both demonstrable and testable.[57] While we assert that the use of algorithms, especially 'black box' algorithms, should be avoided entirely in judicial decision-making,[58] these issues would still need to be addressed even for their use for apparently administrative tasks. This is because even apparently non-judicial functions can profoundly impact judicial outcomes. For example, errors in machine translation of a judgment can mean that lawyers and judges referring to an improperly translated judgment could inadvertently misinterpret it.

Standards of explainability may require the disclosure of a range of information regarding how an algorithm was used to reach a decision, including the role of the human official responsible for the decision and for oversight of the algorithm, The stated goals of its use, the data used to reach the decision; the algorithmic model itself, consisting of its input data, the relative weightage of these inputs, and what training data was used to create it; the potential for error, and the disclosure of the presence or absence of an algorithm itself.[59] These are just some examples of disclosures that may need to be mandated by the regulatory framework to ensure algorithmic accountability for the judiciary, but some such standard of disclosure will be necessary.

Bias

A frequent criticism of the use of algorithms in the judiciary is on the issue of bias. Since humans create these algorithms, they inevitably —and often unconsciously —reflect societal values, biases, and discriminatory practices.[60]

54 Pasquale, Frank. "Toward a fourth law of robotics: Preserving attribution, responsibility, and explainability in an algorithmic society." Ohio St. LJ 78 (2017): 1243.

55 Gilpin, Leilani H., David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. "Explaining explanations: An overview of interpretability of machine learning." In 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA), pp. 80-89. IEEE, 2018.

Gilpin, Leilani H., David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. "Explaining explanations: An approach to evaluating interpretability of machine learning." arXiv preprint arXiv:1806.00069 (2018).

56 That algorithms should be designed and used in compliance with fundamental rights is a well-established principle. See the COE's ethical charter for the use of AI in judicial systems.

57 See 'Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)' for a taxonomy of approaches to designing explainable AI. Also see 'Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models' for approaches specific to deep learning models.

58 Refer to the recommendations in Section 6

59 Diakopoulos, Nicholas. "Accountability in algorithmic decision making." Communications of the ACM 59, no. 2 (2016): 56-62.

60 Internet Society. 2017. 'Artificial Intelligence and Machine Learning: Policy Paper.' Internet Society. Available online at https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper/

The problem of fairness and bias in algorithmic systems have received considerable attention, particularly in the judicial context. The usage of the COMPAS algorithm to predict recidivism in support of sentencing decisions has been at the centre of this controversy once investigations by *Pro Publica* revealed racial bias in the algorithm's predictions.[61]

Bias can cause harm through ostensibly administrative decisions as well. Consider the example of the obvious case for using an algorithm to perform a task currently performed by human officials, the listing of cases. Numerous ethical questions can be raised about algorithms that can easily be raised in this context and how they might affect the outcome of a case. These include concerns regarding the basis of prioritisation, the criteria used to rank these priorities, the likely chances of error and the impacts errors could have, potential biases, and the risks of 'gaming the system' whereby litigants, their lawyers, or other actors choose their actions so as to exploit the rules of the algorithm to gain undue benefits, among others.[62]

Sources of explainability may require disclosure of a range of information regarding how an algorithm was used to reach a decision. These include the role of the human official responsible for the decision and for oversight of the algorithm, and the stated goals of its use, the data used to reach the decision; the algorithmic model itself, consisting of its input data, the relative weightage of these inputs, and what training data was used to create it; the potential for error, and the disclosure of the presence or absence of an algorithm itself.[63] These are just some examples of disclosures that may need to be mandated by the regulatory framework to ensure algorithmic accountability for the judiciary.

Sources of bias are varied, and can creep into an automated process at multiple points. Batya Friedman and Helen Nissenbaum describe three forms of bias that enter computer systems - preexisting bias, which originates in society, institutions, and practices; technical bias, originating in technical features of the system; and emergent bias, which comes into existence only through the use of a system in a given context.[64] Pre existing bias may be sub-categorised as either *societal biases*[65] which are a product of entrenched discrimination and prejudice permeating the dataset used, or biases originating from particular individuals that are influential in the design of the system, such as the designer themselves or a client. Technical bias can come from limitations in the design of computer tools such as how they present information, the use of algorithms that fail to account for the appropriate contextual factors, imperfections in processes designed to be random, and mistakes in translating human actions into algorithmic ones that result in the quantification of qualitative factors, forcing objective categorisation on information and decisions that are actually subjective such as interpretation of law. Finally, emergent bias can result from a system being incompatible

61 -Jeff Larson, Surya Mattu, Lauren Kirchner and Julia Angwin. 2016. 'How We Analyzed the COMPAS Recidivism Algorithm'. ProPublica. May 2016. Available at https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Anupam Chander. 2016. 'The racist algorithm?' Mich. L. Rev., 115, p.1023.Available at https://repository.law.umich.edu/cgi/viewcontent.cgi?article=1657&context=mlr

Avi Feller, Emma Pierson, Sam Corbett-Davies, and Sharad Goel. 2016. A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear. The Washington Post. Available at http://www.cs.yale.edu/homes/jf/Feller.pdf (accessed on 15 July 2019)

62 Diakopoulos, Nicholas. "Algorithmic accountability: Journalistic investigation of computational power structures." Digital journalism 3, no. 3 (2015): 398-415.

62 Algorithmic Decision-Making Based on Machine Learning from Big Data_ Can Transparency Restore Accountability

64 Skitka, Linda J., Kathleen L. Mosier, Mark Burdick, and Bonnie Rosenblatt. "Automation bias and errors: are crews better than individuals?." The International journal of aviation psychology 10, no. 1 (2000): 85-97.

65 Internet Society. 2017. 'Artificial Intelligence and Machine Learning: Policy Paper.' Internet Society. Available online at https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper/

with newly acquired knowledge of a subject or from being inappropriate for the user group it was designed for, either because of an incorrect assumption of their expertise or because of a difference of values between the user group and the designer.

Friedman and Nissenbaum's categories are very useful as a means of thinking about where bias can come from. Still, the complexity and opacity of later machine learning algorithms mean that discerning which of these types of bias exist in a given system or tool is difficult. Many of Friedman and Nissenbaum's categories can appear concurrently. Other useful ways of thinking about bias in this context may frame it in terms of the process - for example, Harini Suresh and John Guttag describe 'historical bias' as a misalignment between the state of the world and the intended values of a given application of machine learning.[66] This corresponds to Friedman and Nissenbaum's 'societal bias' category and occurs before data collection. To take a real-life example, we may not want an algorithm to discriminate on the basis of caste, but since such discrimination exists, it would creep into any data used to both train and use the algorithm.

They then describe 'representation bias' as occurring when the data used to develop and train the algorithm does not represent the population it would be applied to in practice. It would include forms of bias such as 'sampling bias', which results from bias in selecting training data and the non-representativeness resulting from the design of the algorithm itself. This bias occurs during data collection. The most typical form of this is when sampling is thought to be random, and therefore representative, when in fact an unknown factor is causing the sampling process to be non-random and therefore favour some sub-groups over others.[67]

Other forms of bias occur during the preparation of data for model development. Suresh and Guttag describe the problem as follows: "Measurement bias occurs when choosing, collection, or computing features and labels to use in a prediction problem." Recall from Section 2 that supervised learning models utilise labelled data,[68] and learn to classify fresh data from the labels contained in the training data. If there is a bias in the process of assigning these labels, which fails to capture the desired quantities or characteristics adequately, then the model is said to contain measurement bias.[69] It can happen because of variation in data quality between sub-groups in the data, differences between sub-groups in the process of measuring the characteristics used in labelling, or if the process of imposing rigid categorisation upon the data is itself flawed. For example, if the judiciary were to develop scheduling algorithms to list cases, known variations between High Courts in the practice of categorising case types[70] would mean that unless the data is appropriately standardised, using case type as a category for a single algorithm nationwide would result in measurement bias.

The very act of fitting subjective qualities into a model can result in bias. As Ben Green observes, facts processed by machine learning algorithms are reduced to

66 Suresh, Harini, and John V. Guttag. "A framework for understanding unintended consequences of machine learning." arXiv preprint arXiv:1901.10002 (2019).

67 Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. "A survey on bias and fairness in machine learning." arXiv preprint arXiv:1908.09635 (2019).

68 Data is labelled when units of observation have been assigned categories based on a set of characteristics in the data associated with them. These categories are integral to the purpose for which the model is developed and applied.

69 Suresh, Harini, and John V. Guttag. "A framework for understanding unintended consequences of machine learning."

70 DAKSH. 2020 'Deciphering Judicial Data: DAKSH's Database', Damle, Devendra, and Tushar Anand. Problems with the e-Courts data. DAKSH. Available at https://dakshindia.org/wp-content/uploads/2020/08/Case-categorization-paper-FINAL.pdf. No. 20/314. 2020. National Institute for Public Finance and Policy. Available at https://www.nipfp.org.in/media/medialibrary/2020/07/WP_314__2020.pdf

quantitative parameters, leading to a bias towards giving them more importance than important qualitative factors which otherwise would balance out the quantifiable inputs to the decision.[71] He states that "Making decisions via machine learning can therefore distort the values inherent to the task at hand by granting undue weight to quantified considerations at the expense of unquantified ones."[72]

Aggregation bias results from inappropriate combinations of heterogeneous groups, where developing a single model for all sub-groups may be inappropriate for some or all subgroups due to heterogeneity in how the input characteristics relate to the inference sought from the model.[73] To continue with our example of a scheduling algorithm for the judiciary, if cases of a similar type, say cheque bounce cases under S. 138 of the Negotiable Instruments Act, 1881 take varying amounts of time to dispose of due to the variation between states in practices and procedures, developing and utilising a 'one-size-fits-all' algorithm in all states would result in aggregation bias.

Evaluation bias occurs when the performance benchmarks used to evaluate the performance of a machine learning model do not represent the group that the model will be applied to.[74] In other words, it occurs when the criteria used to evaluate the model's performance do not suit the designers' intent.

Deployment bias occurs when bias arises in the course of the use of the machine learning model. It results from a 'mismatch between the problem a model is intended to solve and how it is actually used.'[75] One form of this results from individuals' malicious and targeted efforts, such as in the use of 'adversarial examples'.[76] These are data inputs deliberately designed to fool an algorithm, in order to obtain an undue advantage - colloquially referred to as 'gaming the system'. Many examples of this form of bias are apparent in the use of risk assessment tools in the USA. Even in using machine learning algorithms only in an advisory capacity, there are risks. The impulse to follow a computer's recommendation flows from human "automation bias"—the "use of automation as a heuristic replacement for vigilant information".[77] In this context, Danielle Keats Citron states that "There is a possibility that humans have a tendency to endorse analysis resulting from machine learning, which is probabilistic, as a fact."[78]

Other conceptualisations of the sources of bias overlap with those described above, but merit discussion nonetheless. Some that concern both the design and interpretation of models are well understood by statisticians and data scientists but less so by those who may actually act on the basis of algorithmic inferences. The often-quoted saying that 'correlation does not imply causation' means that simply because patterns exist in the variation of two or more quantities does not mean that one exerts a causal influence over the other. The relationship may be reversed, leading to a bias aptly named 'reverse causality'; it may be mutual, where both quantities exert a direct influence over the other, resulting in 'simultaneity bias '; or, a third unidentified variable exerts an influence

71 Ben Green. 2018. "Fair'Risk Assessments: A Precarious Approach for Criminal Justice Reform.' In 5th Workshop on fairness, accountability, and transparency in machine learning.

72 Green, Ben. "'Fair'Risk Assessments: A Precarious Approach for Criminal Justice Reform.".

73 Suresh, Harini, and John V. Guttag. "A framework for understanding unintended consequences of machine learning."

74 Suresh, Harini, and John V. Guttag. "A framework for understanding unintended consequences of machine learning."

75 Suresh, Harini, and John V. Guttag. "A framework for understanding unintended consequences of machine learning."

76 Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. "Explaining and harnessing adversarial examples." arXiv preprint arXiv:1412.6572 (2014).

77 Skitka, Linda J., Kathleen L. Mosier, Mark Burdick, and Bonnie Rosenblatt. "Automation bias and errors: are crews better than individuals?." The International journal of aviation psychology 10, no. 1 (2000): 85-97.

78 Citron, Danielle Keats. "Technological due process." Wash. UL Rev. 85 (2007): 1249.

on all those in the model but has not been accounted for, resulting in 'omitted variable bias'.[79]

A common thread in the discussion of algorithms, especially in the judicial context, is debate on the role of values in the process of design. Omer Tene and Jules Polonetsky distinguish between 'policy-neutral algorithms' which have not been edited to conform to any policy, and 'policy-directed algorithms', which have.[80] They note that policy-neutral algorithms may still derive from, and serve to entrench, deep-rooted social biases despite any claim of objectivity or of having been modified to correct for these biases. Any algorithms written for the judiciary would undoubtedly need to be policy-directed given the need to ensure they are fair. The process of editing the algorithm to be fair must be governed by rules and standards, it must be audited and reviewed according to those standards. The fact that an algorithm has been edited must be disclosed to whoever is subjected to a decision based on them.[81] Policy-oriented algorithms, even ones that are intended to counter bias, create a different form of bias in doing so. Attempting to satisfy criteria for fairness is difficult, as multiple ways of defining and measuring it exist, and attempting to satisfy multiple parameters for fairness can be difficult or even impossible apart from very narrow conditions.[82] If these conditions for fairness are incompatible with one another, attempting to correct some forms of bias may result in other forms – this hints at the fact that there is no objectively 'neutral' or 'bias-free' process or data, which makes regulation difficult. Thus, the use of algorithms as a part of deciding a case is best avoided completely.

Although significant effort has been devoted to engineering fairness into machine learning algorithms,[83] there are fundamental obstacles that must be overcome. As Ben Green states: "No matter how much data and statistics are involved, however, an algorithm can never be truly neutral and free from normative values."[84] There are numerous implicit normative assumptions built into the choice to use an algorithm for a given task.[85] Decisions based on or supported by algorithms, particularly machine learning and deep learning algorithms, may have intentional effects, and incidental effects.[86] The judiciary would need to establish a form of human oversight with requisite authority and legal and technological expertise to oversee algorithms in any given context. [87] Any regulatory framework for algorithms in the judiciary must provide a procedure to investigate and understand the role of unintended consequences of algorithmic decision-making, standards to evaluate these consequences and their justifiability along with constitutional and jurisprudential principles, and rules to hold the appropriate actor accountable. The latter include the human decision-makers themselves who may be judges or other officers; or vendors responsible for creating the algorithm. Setting standards for liability and the distribution of responsibility for harm is a significant regulatory challenge.[88]

79 Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. "A survey on bias and fairness in machine learning."

80 Tene, Omer, and Jules Polonetsky. "Taming the Golem: Challenges of ethical algorithmic decision-making." NCJL & Tech. 19 (2017): 125.

81 Tene, Omer, and Jules Polonetsky. "Taming the Golem: Challenges of ethical algorithmic decision-making." NCJL & Tech. 19 (2017): 125.

82 Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. 2016. "Inherent trade-offs in the fair determination of risk scores." arXiv preprint arXiv:1609.05807; Berk, Richard, Hoda Heidari, Shahin Jabbari, Michael Kearns, and Aaron Roth. 2018 .'Fairness in criminal justice risk assessments: The state of the art.' Sociological Methods & Research: 0049124118782533.

83 See below in the activities of the regulatory authority for examples

84 Green, Ben. "'Fair' Risk Assessments: A Precarious Approach for Criminal Justice Reform." In 5th Workshop on fairness, accountability, and transparency in machine learning. 2018.

85 Binns, Reuben. "Algorithmic accountability and public reason." Philosophy & technology 31, no. 4 (2018): 543-556.

86 Diakopoulos, Nicholas. "Algorithmic accountability: Journalistic investigation of computational power structures." Digital journalism 3, no. 3 (2015): 398-415.

87 The dangers of faulty, biased, or malicious algorithms requires independent oversight - Ben Shneiderman

88 Tutt, Andrew. "An FDA for algorithms." Admin. L. Rev. 69 (2017): 83.

The issue of bias is connected to transparency. Algorithms are also often so complex that even the engineers and designers who have access to the formulae may struggle or fail to predict the outcome and effects of the algorithm's results. [89] This difficulty in understanding how a machine learning model solves a problem, particularly when combined with a vast number of inputs, makes the problem of minimising bias complicated. As a result, it may be difficult to pinpoint the specific data causing the issue to adjust it. If people feel a system is biased, it undermines their confidence in the technology. [90]

All actors, public and private, must prevent and mitigate against discrimination risks in the design, development and application of machine learning technologies. They must also ensure that there are mechanisms providing access to effective remedy against harms resulting from algorithmic systems before deployment and throughout a system's lifecycle.[91]

Security

As the algorithm learns and interacts with its environment, there are many challenges related to its safe deployment. These challenges can stem from unpredictable and harmful behaviour, including the algorithm's indifference to the impact of its actions. One example is the risk of "reward hacking", where the algorithm finds a way of performing a function that might make it easier to reach the goal, but does not correspond with the designer's intent, such as a cleaning robot sweeping dirt under the carpet. By performing its function in ways that the designers did not anticipate the algorithm could create unintended and unsafe effects. This may be because the reward parameter is not a perfect measure of the desired outcome, or perhaps the optimum level of that parameter is not the maximum level. It is also possible that the manner in which the algorithm resolves the problem it is presented with could have unforeseen side effects.

The safety of an algorithmic agent may also be limited by how it learns from its environment. In reinforcement learning, this stems from the so-called exploration/exploitation dilemma. This means an algorithmic agent may depart from a successful strategy of solving a problem to explore other options that could generate a higher payoff.[92]

The ability to manipulate the training data, or exploit the behavior of an algorithmic agent also highlights issues around transparency of the machine learning model. Disclosing detailed information about the training data and the techniques involved may make an algorithmic agent vulnerable to adversarial learning.

Safety and security considerations must be taken into account in the debate around transparency of algorithmic decisions. [93] To give an example from the application of a scheduling algorithm in the judicial context. If the workings of the algorithm are public, it can be deliberately manipulated by misrepresenting the information that it uses to prioritise cases for listing, perhaps through choosing what information to include in petitions and other documents. This could potentially be done in a manner that

89 Adrienne Yapo and Joseph Weiss. 2018. 'Ethical implications of bias in machine learning.' Proceedings of the 51st Hawaii International Conference on System Sciences. Available at https://doi.org/10.24251/HICSS.2018.668

90 Internet Society. 2017. 'Artificial Intelligence and Machine Learning: Policy Paper.' Internet Society. Available online at https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper

91 The Toronto Declaration. Available at https://www.torontodeclaration.org/declaration-text/english/

92 Internet Society. 2017. 'Artificial Intelligence and Machine Learning: Policy Paper.' Internet Society. Available online at https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper/

93 Internet Society. 2017. 'Artificial Intelligence and Machine Learning: Policy Paper.' Internet Society. Available online at https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper/

contradicts the intention of the designers, without sufficient human oversight and technological safeguards.

Accountability

Any regulatory system governing the use of algorithms in the judiciary must provide for accountability in the event that the use of algorithms results in the violation of the rights of individuals or groups. Such a regulatory system must specify how accountability is to be fixed in the event algorithm-assisted decisions violate due process and fair trial requirements. The GDPR provides the right "not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her" unless certain safeguards are met.[94].

However, the issue with fixing accountability for harmful decisions is that it presumes a level of transparency in the algorithm. An essential feature of learning algorithms is their ability to generate rules without step-by–step instructions. While the technique helps the algorithm perform complex tasks such as face recognition or interpreting natural language, it gives programmers less control. Unlike in non-machine learning algorithms, where the reasoning behind an algorithm's specific output can often be explained, this is not true for machine learning algorithms. It is difficult to fix accountability when one is not able to explain why a specific action was taken. The opacity of the reasoning behind an algorithm's actions complicates the already difficult question of software liability. It is necessary to clarify how and when the manufacturer, operator, and the programmer will be held liable [95] This is discussed in more detail in Section 6.

Accessibility and Inclusion

Anyone with a disability should be treated with human dignity and be included in the enjoyment of fundamental human rights. Companies and developers can use inherent respect for human dignity and human rights to act on anticipated harms. By proactively addressing negative impacts, as in the case of accessibility for people with disabilities, developers can take steps to advance human rights. [96]

Like all technologies before it, algorithms will reflect the values of their creators. A principle of inclusion should underlie the design of these alogrithms to ensure that a range of ethical perspectives is heard. Otherwise, we risk constructing machine intelligence that mirrors a narrow and privileged vision of society, with its old, familiar biases and stereotypes.[97]

Privacy

Developments in algorithms and the proliferation of 'big data' allow the re-identification of people,[98] and more personally identifiable information can be added to this to create complex profiles of people and make predictions about their lives.[99] They render conventional regulatory means of preserving privacy inadequate to deal with emergent risks. For example, redaction or anonymisation of data has been shown to be ineffective and impermanent, and is easily subverted with the use of machine learning algorithms to reconstruct a person's identity.[100]

94 Brian Sheppard. 2018. 'Warming up to inscrutability: How technology could challenge our concept of law.' University of Toronto Law Journal 68(supplement 1): 36-62.

95 Internet Society. 2017. 'Artificial Intelligence and Machine Learning: Policy Paper.' Internet Society. Available online at https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper/

96 Governing Artificial Intelligence

97 Adrienne Yapo and Joseph Weiss, Ethical Implications Of Bias In Machine Learning, 5369

98 Tamò-Larrieux, Aurelia, Tamò-Larrieux, and Seyfried. Designing for privacy and its legal framework. Cham: Springer, 2018.

99 Crawford, Kate, and Jason Schultz. "Big data and due process: Toward a framework to redress predictive privacy harms." BCL Rev. 55 (2014): 93.

100 Paul Ohm. 2009. 'Broken promises of privacy: Responding to the surprising failure of anonymization,' UCLA Law Review, 5: 1701; Vincent Toubiana and Helen Nissenbaum. 2011. 'An analysis of google log retention policies'

In a particularly notable and alarming example, researchers were able to write an algorithm that could guess US citizens' Social Security Numbers, an identification which exposes them to significant fraud and other kinds of vulnerability, using publicly available data.[101] Awareness of the considerable scope for harm resulting from machine learning with big data has grown due to events such as the Cambridge Analytica-Facebook scandal, in which complex psychological profiles were used to target Facebook users with ads on the basis of their predicted political leanings.[102]

Explainability has relevance for privacy. The inferences that algorithms can draw from data are often unpredictable, and bear little correlation to intuition.[103] Privacy regulations often incorporate consent as a precondition for legal and legitimate processing of data. The question that emerges is whether consent can be considered valid when a person cannot reliably predict what an algorithm will infer about them. This is particularly troubling if such an algorithm is used to make a decision concerning them. A person may become the subject of a decision based on an algorithmic inference, even if their own data was not used to draw the initial inference thanks to the deductive capabilities of machine learning algorithms.[104] To use an

101 Alessandro Acquisti and Ralph Gross. 2009. 'Predicting social security numbers from public data.' Proceedings of the National Academy of Sciences, 106 (27): 10975-10980.

102 Carole Cadwalladr and Emma Graham-Harrison. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

103 Andrew D. Selbst and Solon Barocas. 2018. 'The intuitive appeal of explainable machines.' Fordham L. Rev. 87: 1085.

104 Martijn Van Otterlo. 2013. 'A machine learning view on profiling.' Privacy, Due Process and the Computational Turn-Philosophers of Law Meet Philosophers of Technology. Abingdon: Routledge: pp. 41-64.

example given by Martijn Van Otterlo,[105] an algorithm on an E-Commerce website can observe that people who buy a particular product are more likely to buy another, and use this inference to recommend products to new users, among whom this pattern has not yet been observed.

Many proposed policy frameworks for India, including NITI Aayog's papers on strategy and accountability,[106] have rightfully emphasised the urgency of adopting a privacy law to enable responsible use of AI. However, they do not address numerous shortcomings of the current proposed privacy bills. India has a data protection bill for personal data[107] under consideration by a Joint Parliamentary Committee, the Personal Data Protection Bill, 2019[108]. This Bill is a modified version of a bill drafted by the Justice B.N. Srikrishna Committee.

Crucially, the bill exempts the judiciary from many of its key provisions,[109] including the rights of people to whom data pertains, called data principals;[110] and most of the obligations of those who acquire and process their personal data, called data fiduciaries.[111] However, the exemptions do not apply to its administrative actions.[112] The Bill's definition of harm,[113] which is relevant for AI,[114] applies to judicial data. However, this definition of harm is not connected to the misuse of data in the bill's provisions, but is tied to the obligations of fiduciaries and penalties for misconduct.

The Bill imposes an obligation to undertake a personal data impact assessment before the use of new technologies, which would presumably include advanced learning algorithms, or genetic or biometric data.[115] It also contains obligations to maintain accurate, current, and complete data, but does not address concerns specific to AI such as bias and representativeness. [116] The judiciary is notably exempt from this provision, too. [117] There are no obligations to inform a principal about automated data processing and no right to opt-out from it. [118]

Judicial data contains a host of information that renders people vulnerable.[119] In the absence of a comprehensive privacy framework for the judiciary, the exemptions in the already limited provisions of the Bill mean that any harm caused by algorithmic decision making within the judiciary can only be redressed through interpretation of the Constitution and other laws.

105 Martijn Van Otterlo. 2013. 'A machine learning view on profiling.' Privacy, Due Process and the Computational Turn-Philosophers of Law Meet Philosophers of Technology. Abingdon: Routledge: pp. 41-64.

106 NITI Aayog. 2018. National Strategy On Artificial Intelligence. https://niti.gov.in/sites/default/files/2019-01/NationalStrategy-for-AI-Discussion-Paper.pdf, NITI Aayog. 2020. Working Document on Responsible AI for All https://niti.gov.in/sites/default/files/2020-07/Responsible-AI.pdf

107 "personal data" means data about or relating to a natural person who is directly or indirectly identifiable, having regard to any characteristic, trait, attribute or any other feature of the identity of such natural person, whether online or offline, or any combination of such features with any other information, and shall include any inference drawn from such data for the purpose of profiling;" S.3(28), PDPB 2019

108 http://loksabhaph.nic.in/Committee/CommitteeInformation.aspx?comm_code=73&tab=1 As of 3 October 2020

109 S. 36 (c), Personal Data Protection Bill (PDPB), 2019

110 S. 3(14) , Personal Data Protection Bill (PDPB), 2019

111 S. 2(13) , Personal Data Protection Bill (PDPB), 2019

112 An Analysis of 'Harm' defined under the draft Personal Data Protection Bill, 2018 https://www.dvara.com/blog/2019/10/29/an-analysis-of-harm-defined-under-the-draft-personal-data-protection-bill-2018/

113 Section 3(20) of the PDP Bill 2019: "Harm includes – (i) bodily or mental injury; (ii) loss, distortion or theft of identity; (iii) financial loss or loss of property; (iv) loss of reputation or humiliation; (v) loss of employment; (vi) any discriminatory treatment; (vii) any subjection to blackmail or extortion; (viii) any denial or withdrawal of a service, benefit or good resulting from an evaluative decision about the data principal; (ix) any restriction placed or suffered directly or indirectly on speech, movement or any other action arising out of a fear or being observed or surveilled; or (x) any observation or surveillance that is not reasonably expected by the data principal.

114 Amber Sinha and Elonnai Hickok. 2018. 'The Srikrishna Committee Data Protection Bill and Artificial Intelligence in India'. Centre for Internet and Society.https://cis-india.org/internet-governance/blog/the-srikrishna-committee-data-protection-bill-and-artificial-intelligence-in-india.

115 Amber Sinha and Elonnai Hickok. 2018. 'The Srikrishna Committee Data Protection Bill and Artificial Intelligence in India', S. 27(1) PDPB 2019,

116 Amber Sinha and Elonnai Hickok. 2018. 'The Srikrishna Committee Data Protection Bill and Artificial Intelligence in India', S. 28(1) PDPB 2019,

117 S. 36(c) , PDPB 2019,

118 Amber Sinha and Elonnai Hickok. 2018. 'The Srikrishna Committee Data Protection Bill and Artificial Intelligence in India'

119 <Refer JDP Paper>

There are four situations in which privacy concerns arise:

1. Information collection operations, which, in a judicial context, refer to information about the parties, legal professionals and the third parties connected to the proceedings.[120]
2. Information processing, referring to the procedures involved in using, storing, and manipulating data. From a judicial perspective, it is important to highlight two modes of information processing. One of them is secondary use: data handled for one purpose might be used for other ends, thereby frustrating the expectations that legitimated the original use. The second is the matter of exclusion, where persons might be deprived of the possibility of exercising their privacy rights because they lack adequate information about the existence and nature of the information operations which affect them. In both cases, there is an intrusion into a person's private life, both because the unexpected uses or lack of information affect that person's right to informational self-determination and because the omitted or reused information might constrain their decisions, for example by preventing them from seeking recourse for an automated judicial decision. [121]
3. Information dissemination may also interfere with a person's privacy, as it might allow access to information that the person would not want to share, to specific targets or the public in general. In judicial automation, the impact of information dissemination is particularly relevant when one considers the automation of proceedings involving confidential information, such as those concerned with children's rights.
4. Some forms of processing of data can intrude on a person's liberty to make decisions regarding their private life. Enforcing decisional privacy in the context of increased use of algorithmic technologies for administrative functions in the judiciary requires that one understands how automation tools may empower citizens rather than unduly restricting their decisional privacy.

Right to a human decision

Article 22 of the GDPR endows natural individuals with "the right not to be subject to a decision based solely on automated processing including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." On its face, it fashions an opt-out of automated decision-making.[122] Automated decision making is a process through which a duly programmed IT system can produce a significant decision for the subjects involved based exclusively on the algorithmic evaluation of the personal data of the profiled subjects/users without the aid of human intervention. [123]

Most systems of organisation consist of rules, and a decision-making system that is merely about obeying rules might be replaced by software quite easily. But real systems of human ordering and decision-making, even those based on rules, are not that simple. Disputes can be "easy cases"- those covered by settled rules or "hardcases" in which the application of rules is not striaghtforward, or where the rules contradict each other. Fair results in hard cases still depend on accessing something that remains, for now,

120 Marco Almada and Maria Dymitruk, Privacy and Data Protection Constraints to Automated Decision-Making in the Judiciary, 23

121 Marco Almada and Maria Dymitruk, Privacy and Data Protection Constraints to Automated Decision-Making in the Judiciary, 25

122 Aziz Z.Huq. 2020. 'A Right to a Human Decision.' Virginia Law Review. 106: 611.

123 Elena Falletti. 2019. 'Automated Decisions and Article No. 22 GDPR of the European Union: An Analysis of the Right to an "Explanation"' Available at SSRN 3510084. 6

human, whether we call it moral reasoning, a sensitivity to evolving norms, or a pragmatic assessment of what works. [124]

However, the automation of routine procedure might help produce both a much faster legal system and also free up the scarce resource of highly trained human judgment to adjudicate the hard cases, or to determine which are the hard cases. The judiciary's mental resources are squandered on thousands of routine matters; there is promise in a system that leaves judges to do what they do best: exercising judgment in the individual case, and humanising and improving the written rules.[125]

Related to this right is the right of the data subject to express their opinion. The data subject, to use the GDPR's term, should be allowed to express their point of view prior to the use of an algorithm in a context which would impact them. The data controller should deploy measures to prevent a situation where the subject of a final decision is not consulted before the decision is taken. The other measure to safeguard the data subject's rights, freedoms and legitimate interests is the right to contest the decision. Fulfilling this right in the context of automated judicial decision-making entails primarily appealing against the automated decision taken in the course of court proceedings.[126]

124 Tim Wu, Will Artificial Intelligence Eat The Law? The Rise Of Hybrid Social-Ordering Systems, 2003

125 Tim Wu. 2019. 'Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems.' Columbia Law Review 119(7): 2001-2028.

126 Marco Almada and Maria Dymitruk, Privacy and Data Protection Constraints to Automated Decision-Making in the Judiciary, 22

Section 5:

International Experience

The debate over the usage and ethics of algorithms and artificial intelligence in the judiciary is an important topic of discussion in many international jurisdictions. Some jurisdictions favour a generous approach to use algorithmic tools in a bid to increase efficiency, while others caution that the inclusion of advanced technology tools like artificial intelligence within the judiciary should be measured. This section will provide an overview of how some jurisdictions are engaging with algorithms and artificial intelligence within their respective judicial landscapes.

UNITED KINGDOM

The United Kingdom (UK) is currently initiating research on the use of algorithms and artificial intelligence in the judiciary. There is considerable funding allocated to examine the impact of algorithms in the judiciary. The UKRI (UK Research and Innovation) allocated a proportion of their funding to the Alan Turing Institute to work on criminal reforms using artificial intelligence.[127] The UK has identified that regulatory issues are important in allowing the use of algorithms and artificial intelligence in the judiciary. The Law Tech Delivery Panel has been established by the Lord Chancellor to identify the various barriers to implementing artificial intelligence.[128]

In the UK, more than ten authorities are looking into artificial intelligence regulatory aspects. No specific rules govern the use of artificial intelligence by the legal sector.[129] The role of the Solicitors Regulation Authority (SRA) is of particular importance. The government has allocated funding of £ 700,000 to the SRA to incorporate artificial intelligence in the legal sector.[130] As per the SRA guidelines, legal service providers incorporating artificial intelligence in their services should adequately inform their clients, to ensure that client expectations and service ethics are maintained.[131]. The principles governing legal artificial intelligence in the UK are intertwined with the data usage policies.[132]

The UK has approached the issue of incorporating artificial intelligence in the judiciary through conceptual research. Such research will inform frameworks for the legal engagement of artificial intelligence in the UK. The most prominent ones are at the Oxford Internet Institute, studying the constitutional implications of artificial intelligence in the judiciary. Such projects ensure a rigorous principle-based approach to the use of artificial intelligence in the judiciary.

EUROPEAN UNION (EU)

The EU High-Level Expert Group on Artificial Intelligence highlighted two factors to be considered when deciding on the use of artificial intelligence in any process[133]

1. Deciding on the usefulness of a given artificial intelligence application, rather than focusing on efficiency or availability;
2. Ensuring the participation of stakeholders (Including marginalised communities and others who will face the implications of using artificial intelligence).

127 The Legal Education Foundation. 2019. Digital Justice : HMCTS data strategy and delivering access to justice. United Kingdom, p. 34. Available online at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/835778/DigitalJusticeFINAL.PDF (accessed on 21 May 2021).

128 The Legal Education Foundation, Digital Justice : HMCTS data strategy and delivering access to justice.

129 Richard Kemp. 2018. 'Legal Aspect of Artificial Intelligence', Kemp IT Law, September, available online https://www.kempitlaw.com/legal-aspects-of-artificial-intelligence-3/ (accessed on 21 May 2021).

130 Ministry of Justice, Legal Support: The Way Ahead.

131 Kemp, 'Legal Aspect of Artificial Intelligence'.

132 Government of UK. 2018. 'Data Ethics Framework', GOV.UK, June 13, available online at https://www.gov.uk/government/publications/data-ethics-framework (accessed on 21 May 2021).

133 Michael Veale. 2020. 'A Critical Take on the Policy Recommendations of the EU High-Level Expert Group on Artificial Intelligence', SSRN, January 26, available online at https://ssrn.com/abstract=3475449 (accessed on 21 May 2021).

COUNCIL OF EUROPE

The European Commission for the Efficiency of Justice (CEPEJ) has adopted a charter establishing principles for the use of what they describe as "AI" in justice systems.[134] It suggests the usage of artificial intelligence in civil, commercial and administrative cases, provided that there is an option of appeal available, pursuant to such artificial intelligence usage. Some EU jurisdictions are experimenting with simple algorithms in online dispute resolution for low-value disputes.[135]

The CEPEJ charter lists the following principles that ought not to be violated in civil, commercial, and administrative proceedings:

RIGHT OF ACCESS TO A COURT:

This refers specifically to the right to a fair trial as guaranteed by the European Convention on Human Rights

ADVERSARIAL PRINCIPLE:

The qualitative and quantitative data used to calculate the claim should be accessible to all, including parties to a dispute.

EQUALITY OF ARMS:

Certain technological advantages benefit some litigants more, and there should be checks and balances against this.

INDEPENDENCE OF JUDICIARY:

There should not be total reliance on the automated resolution of disputes, decisions etc., and the judiciary should maintain checks and balances.

RIGHT TO COUNSEL:

Some techonology-enabled online dispute resolution tools are based on the premise of reducing the need for legal representation,[136] but the legal representation should be provided for if parties require it.

From the perspective of criminal reforms and technology, it is suggested that algorithmic tools be used to accelerate trials by condensing information for the judges.[137] The CEPEJ Charter exercises caution while initiating the usage of tools like predictive policing within the criminal justice framework given the impact on the right to liberty.[138] The charter calls for robust monitoring of any initiation of such tools by judicial stakeholders and reiterates that such a decision will not always be about efficiency.[139]

A proposed checklist before initiating algorithms and artificial intelligence in the

134 European Commission for the Efficiency of Justice (CEPEJ). 2018. European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, available online at https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (accessed on 21 May 2021).

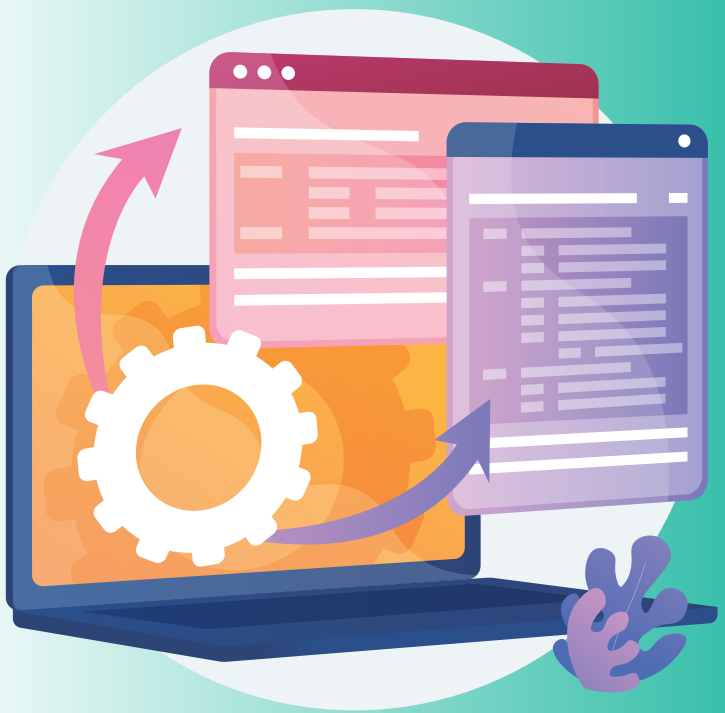
135 European Commission for the Efficiency of Justice (CEPEJ). European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, p. 44

136 John Sorabji. 2017.'The online solutions court – a multi-door courthouse for the 21st century', Civil Justice Quarterly 36 (1) : 86.

137 European Commission for the Efficiency of Justice (CEPEJ). European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment.

138 European Commission for the Efficiency of Justice (CEPEJ). European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, pp.53-54

139 European Commission for the Efficiency of Justice (CEPEJ). European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, pp.53-54



legal framework:[140]

1. Setting up of a cyber-ethics framework that respects fundamental rights before initiating the usage of artificial intelligence;
2. Public debates on the tools, which include engaging with all stakeholders in the judiciary and entities that design the tools;
3. Constant monitoring of any new algorithmic and artificial intelligence tools used;
4. Regulating big data and its usage, including within the judiciary.

The proposed use cases for artificial intelligence in the judiciary as per the CEPEJ charter:

The charter has proposed a phase-wise introduction of artificial intelligence in the judiciary. The first phase is focused on improving access to laws and using tools to improve judicial functioning. The first three use cases recommended for using artificial intelligence are:

1. Case law enhancement, which involves using natural language processing to make case law (as well as legislation, conventions, scholarship, and regulations) more easily searchable;
2. Improving the access to laws through court forms and chatbots for efficiency and access for citizens;
3. Creating strategic tools to process and understand court analytics and data to suggest improvements in court functioning based on advanced performance metrics.

The second phase of using artificial intelligence involves its use in calculating claim amounts, supporting alternative dispute resolution mechanisms and helping in criminal investigations. The EU suggests extra caution and studies before initiating artificial intelligence in judge profiling, making judicial decisions and criminal proceedings.

CHINA

The Chinese government has been particularly keen to deploy algorithms and artificial intelligence to improve transparency and accountability in the judiciary.[141] China has tested the implementation of some artificial intelligence tools in its various 'internet courts.' The first of such courts were launched in 2019 in Hangzhou, Beijing, and Guangzhou. In the internet courts, cases involving internet services are usually conducted entirely online. Interestingly, these internet courts have been used as pilots for new technologies, which are then gradually incorporated into the workings of all courts. For example, the e-filing processes piloted in the internet courts have now been applied by court divisions in Shanghai, Binhai, and Shenzen.[142] In Shanghai, a six-month pilot project launched in 2020 aims to use algorithms and artificial intelligence to automate some of the work done by courtroom clerks. Ten courts have introduced artificial intelligence tools to perform tasks such as the transcription of case notes, retrieval of files, and management of digital evidence.[143] The changes are expected to free up more time for clerks to work on trial preparations,

140 European Commission for the Efficiency of Justice (CEPEJ). European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, pp.59,60

141 Huw Roberts, Josh Cowls, Jessica Morley, Mariarosaria Taddeo, Vincent Wang, and Luciano Floridi. 2021. 'The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation', AI & SOCIETY 36(1): 59-77. available online at https://doi.org/10.1007/s00146-020-00992-2 (accessed on 21 May 2021).

142 Guodong Du and Meng Yu. 2019. 'China's Supreme Court Issues a White Paper on Chinese Courts'. China Justice Observer, available online at https://www.chinajusticeobserver.com/a/supreme-peoples-court-issues-a-white-paper-on-china-court-and-internet-judiciary (accessed on 21 May 2021).

143 S. Dai. 2020. 'Shanghai judicial courts start to replace clerks with AI assistants', Retrieved from South China Morning Post, 1 April, available online at https://www.scmp.com/tech/innovation/article/3077979/shanghai-judicial-courts-start-replace-clerks-ai-assistants (accessed on 21 May 2021).

as opposed to performing administrative functions.[144] This project was launched following an earlier instruction in 2017 from the Supreme People's Court that all courts were required, as far as possible, to use speech recognition software in generating legal transcripts.[145] The Chinese model emphasises the need for pilot projects before incorporating the changes in technology to all the other courts.

Some of the reforms in using artificial intelligence has focussed on compelling judges to be clear in their reasoning and follow precedent. Two major technologies have been used to enable this. Firstly, there is a 'similar case' identification system, which uses artificial intelligence to help judges identify cases with similar factual backgrounds for their reference. This was adopted by several courts, and the Hainan High People's Court has urged its use at all levels of the regional judiciary. The second kind of technology detects 'abnormal' judgements, alerting senior judges that a decision has deviated from past cases and may have been improperly influenced.[146] Notably, the implementation of these protocols is directly led by the Supreme People's Court.[147] The two guiding ethical principles for artificial intelligence in the courts in China seem to be: (1) It must be conducive to fairness, for example by reducing delays and improving accessibility, and (2) it must promote transparency and accountability.

NETHERLANDS

In the Netherlands, the use of artificial intelligence and algorithms is a part of the larger conversation on the 'Dutch Digitisation Strategy'.[148] The aim is to transform administration by introducing innovation through big data analytics and artificial intelligence in various fields.[149] A section of the strategy is dedicated to protecting the fundamental rights of the people and public values that can be easily disrupted by new technologies.[150] The government has formed inter-departmental groups and commissioned independent research related to artificial intelligence. For example, one ministry set up a transparency lab that issues guidelines for transparency in the use of algorithms.[151]

The Netherlands will soon begin using artificial intelligence in its judicial system. The Ministry of Justice has organised roundtable discussions on the subject,[152] and pilot projects have begun in some courts. For example, in the District Court of East Brabant, an ongoing project in collaboration with three universities seeks to develop a system to deploy artificial intelligence in resolving traffic violation cases. The appeals process for such violations will be partly automated.[153] The study also aims to develop a case management tool for judges to use in handling such cases.

FRANCE

France has developed a national artificial intelligence strategy that focusses on four key areas: health, transport, environmental

144 Dai, 'Shanghai judicial courts start to replace clerks with AI assistants'.

145 Dai, 'Shanghai judicial courts start to replace clerks with AI assistants'.

146 Roberts et. al.. 'The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation'

147 Du and Yu. 'China's Supreme Court Issues a White Paper on Chinese Courts'

148 Central Government of Netherland. 2018. Dutch Digital Strategy. Netherland: Central Government of Netherland. Available online at https://www.rijksoverheid.nl/documenten/rapporten/2018/06/01/nederlandse-digitaliseringsstrategie (accessed on 21 May 2021).

149 European Commission. 2019. Netherlands AI Strategy Report. EU: European Commission. Available online at https://ec.europa.eu/knowledge4policy/ai-watch/netherlands-ai-strategy-report_en (accessed on 21 May 2021).

150 Central Government of Netherland. Dutch Digital Strategy. .

151 The Netherlands. 2019. Strategic Action Plan for Artificial Intelligence, available online at https://www.government.nl/documents/reports/2019/10/09/strategic-action-plan-for-artificial-intelligence

152 Gijs Van Til, 2019. 'Report: Automating Society – Netherlands'. Algorithm Watch. 28 January. Available online at https://algorithmwatch.org/en/automating-society-netherlands/ (accessed on 21 May 2021).

153 A. D. Reiling. 2020. Courts and Artificial Intelligence. International Journal for Court Administration, 11(2):8.

affairs, and defence and security systems.[154] It released a national strategy in 2018. However, the government has been reluctant to use AI in the judiciary. In 2019, France effectively banned the use of predictive analytics in the legal system.

In recent years, there have been debates in France as to whether judges' names should be redacted on judgements when they are published online. It is reported that the ban on data analytics may have been a compromise solution which allowed unredacted judgements to be in the public domain, but with restrictions on their use.[155]

CANADA

In 2017, the Government of Canada appointed the Canadian Institute for Advanced Research (CIFAR) as the body responsible for leading the country's $125 million 'Pan-Canadian Artificial Intelligence Strategy.'[156] At the federal level, efforts to increase the uptake of artificial intelligence in the legal system are spearheaded by the Department of Justice. The Department's 'Artificial Intelligence (AI) Task Force' is tasked with examining opportunities for the legal sector in the artificial intelligence space. It is also in charge of handling pilot projects involving new technologies.[157] The task force comprises academics, civil society groups, industry representatives, and government leaders.[158]

The Privacy Act (PA) and The Personal Information Protection and Electronic Documents Act (PIPEDA) regulate the government and private organisation's handling of personal information. Under Canadian law, 'personal information' is information that could, either by itself or in conjunction with other information, be used to identify an individual.[159] Under PIPEDA, commercial businesses must adhere to ten core principles when collecting or processing users' personal information. Data management policies must also be publicised.[160]

Several initiatives in Canada have sought to foster ethical artificial intelligence

154 European Commission. 2020. Knowledge for policy: France AI Strategy Report. 5 August. Available online at https://ec.europa.eu/knowledge4policy/ai-watch/france-ai-strategy report_en (accessed on 21 May 2021).

155 Tim Zubizarreta. 2019. New France law bans use of analytics to determine judge behavior, Jurist, 5 June. Available online at https://www.jurist.org/news/2019/06/new-france-law-bans-use-of-analytics-to-determine-judge-behavior/ (accessed on 20 August)

Carl Schonander.2019. French judicial analytics ban undermines rule of law, CIO, 3 July https://www.cio.com/article/3406797/french-judicial-analytics-ban-undermines-rule-of-law.html

156 Canadian Institute for Advanced Research. 2020. CIFAR Pan-Canadian Artificial Intelligence Strategy, CIFAR, available online at https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy (accessed on 21 May 2021).

157 Government of Canada, 2018. Department of Justice: 2018-19 Departmental Plan. Government of Canada, April. Available online at https://www.justice.gc.ca/eng/rp-pr/cp-pm/rpp/2018_2019/rep-rap/p3.html (accessed on 21 May 2021).

158 Government of Canada. 2019. Government of Canada creates Advisory Council on Artificial Intelligence, Government of Canada, July. Available online at https://www.canada.ca/en/innovation-science-economic- (accessed on 21 May 2021).

159 Office of the Privacy Commissioner of Canada. 2018. What is personal information? Office of the Privacy Commissioner of Canada, Januanry. Available online at https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/02_05_d_15/#heading-0-0-1 (accessed on 21 May 2021).

160 Government of Canada. 2020. Personal Information Protection and Electronic Documents Act, Justice Laws, July. Available online at https://laws-lois.justice.gc.ca/ENG/ACTS/P-8.6/index.html (accessed on 21 May 2021).

development. The 'Chief Information Officers Strategy Council' (CIOSC) in Canada produces detailed product standards and forms technical committees to discuss developments in the justice space. Some provincial governments in Canada sought to decongest the legal system by developing 'Online Dispute Resolution' (ODR) platforms for low-value civil claims. British Columbia's online 'Civil Resolution Tribunal' offers a 'Solutions Provider' to classify claims. However, the actual decision-making process requires human mediators.[161] The 'Solutions Provider' utilises basic algorithms to form 'pathways' for users to access helpful legal information. The platform can also be useful in other ways. In 'strata disputes', which generally involve two parties from the same condominium or residents' association, the CRT can help users generate a template letter that they can send to their condominium's councils.[162] The CRT was developed and funded by the government of British Columbia.

"Platform to Assist in the Resolution of Litigation Electronically" (PARLe) is an ODR platform operating in Quebec. The project is the result of a partnership between the non-profit 'Cyberjustice Lab' and Quebec's Ministry of Justice and consumer protection agency (the OPC).[163] PARLe uses a 'chatbot' and an artificial intelligence-supported document triage system to determine whether the applicant is eligible to use the service for their claim.[164] This enables the system to gain information about the types of documents associated with small-claims cases. The PARLe platform is open-source and customisable for the needs of particular sectors.[165] Open-source designs allow for the needs of each context and jurisdiction to be appropriately recognised whilst also potentially allowing the artificial intelligence tools to be enhanced and iterated upon based on their performance in different legal environments.

161 Civil Resolution Tribunal. 2020. Starting a Dispute. Civil Resolution Tribunal, Available online at https://civilresolutionbc.ca/tribunal-process/starting-a-dispute/#1-apply-from-the-solution-explorer (accessed on 28 August)

162 Shannon Salter. 2017. Online Dispute Resolution and Justice System Integration: British Columbia's Civil Resolution Tribunal. Windsor Y B Access, 34(1), 112-139.

163 Nicolas Vermeys and Karim Benyekhlef. 2017. Publicly Funded Consumer ODR Is Now a Reality in Quebec, Slaw, February. Available online at http://www.slaw.ca/2017/02/10/publicly-funded-consumer-odr-is-now-a-reality-in-quebec/ (accessed on 21 May 2021).

164 PARLe. 2019. Transform the Court Experience with Online Dispute Resolution. Cyberjustice Laboratory. Available online at https://cyberjustice.openum.ca/files/sites/102/Livret_LABOCJ_PARLe_demilettre_GN-1-Corrige%CC%81-2.pdf (accessed on 21 May 2021).

165 PARLe. Transform the Court Experience with Online Dispute Resolution.

Section 6:

An institutional arrangement for the regulation of algorithms

After describing the regulatory challenges posed by recent advances in algorithms and the values that should be preserved in their use, we propose a regulatory framework for algorithmic accountability in the justice system and the legal profession.

The necessity of a regulatory body under the judiciary

JUDICIAL INDEPENDENCE

Many of the ethical concerns that algorithmic decision-making raise vary between sectors, in both composition and magnitude. Many recommend that the regulation of advanced algorithmic technologies, typically grouped under the umbrella term of 'AI', be taken up by sectoral regulators. Notably, NITI Aayog recommended this in their approach paper on an AI strategy for India[166] and its subsequent working document, 'A Responsible AI for All.'[167] For ethical concerns that cut across sectors, others recommend relying on a central regulator to frame an agenda, drive research, and frame broader policy and ethical standards. The roles of fully-centralised regulators include developing standards, tests, audit tools and methodologies, and certifications for technologiesSuch a regulator would lack the judicial expertise and authority to make decisions that are binding on the judiciary. They might also violate principles of judicial independence. Therefore, independent regulation of algorithms in a judicial context is necessary.

An independent regulatory body for algorithms in the judiciary due to their unique regulatory challenges. Both legal and technical expertise are necessary to understand the implications of using a particular algorithm in a given context. Regulation may be more successful if overseen by a body accountable to the judiciary than by judges and court staff themselves, who may lack the technical expertise required for this task.

ADVANCED ALGORITHMS CHALLENGE LEGAL CONCEPTS

There are several unique features of algorithms that challenge fundamental legal concepts. The most important of these is the fact that algorithms can make choices. Sometimes, these choices have a moral dimension, even if it is not readily apparent, because the exercise of morality in decision-making is often implicit and assumed during the design and development process. The challenge, according to Jacob Turner, is that technology is 'interposing itself between humans and an eventual outcome', in matters that would be said to involve the exercise of moral judgment if done by a human.[168] In the judicial context, where the impacts of even mundane, ostensibly administrative choices can have serious consequences for those involved, this is even more concerning.

As discussed earlier, applying rules to algorithms intended to be administered by humans, even administrative and procedural ones, may become problematic. This is because algorithms typically require inputs to be in quantitative/numerical form, but not all factors can easily be measured in a quantitative/numerical - to do so may require us to make certain assumptions. To return to the example of a scheduling algorithm, the characteristics of cases that determine their duration may include factors such as the details of events that led to a dispute such as the personal motivations of the litigants in contesting or prolonging the case. These are not easily captured in an objective scheme of classification, but the developers may choose to assume that the laws under which the case is filed may capture it adequately. The

166 NITI Aayog. 2018. National Strategy on Artificial Intelligence

167 NITI Aayog. 2020. 'Working Document on Responsible AI for All'. NITI Aayog. Available at https://niti.gov.in/sites/default/files/2020-08/Responsible_AI_05082020.pdf

168 Turner. Robot rules: Regulating artificial intelligence.

capability of machine learning algorithms to learn in a manner unintended by their creators, and eventually, to write new algorithms that give them complex capabilities is also problematic.[169]

Turner argues that they are qualitatively different from other entities capable of independent adaption, like bacteria in a lab, because of their capacity to interact with and follow rules that humans have written.[170] This fact means that it is tempting to utilise algorithms in domains such as law because their immense capabilities could potentially be applied to solve numerous problems. However, Turner identifies three features of the existing law that make them inadequate to govern algorithms appropriately: the law leaves gaps for human discretion; it is not static and may be overruled by courts; and the increasing opacity and unpredictability of complex algorithmic decisions, even in compliance with the law, is problematic because the law as designed to be followed by humans assumes a degree of moral capacity in those humans, and even trivial decisions very often have a moral component.[171] With regard to the latter point, he states that "the increasing unpredictability of AI renders it ever more difficult to tether each decision AI takes to humans through a traditional chain of causation."[172]

ADVANCED ALGORITHMS PRESENT SIGNIFICANT PROBLEMS SPECIFIC TO THE JUDICIAL CONTEXT

The principles of natural justice, as ruled in *A. K. Kraipak & Ors. Etc vs Union Of India,*[173] are that no one shall be a judge in his own case, no decision shall be given against a party without affording him a reasonable hearing, and judicial enquiries must be held in good faith and without bias, and not arbitrarily or unreasonably. Algorithms may violate these principles, given the various forms of bias that may occur and the difficulty of understanding their reasoning, as discussed earlier.

While Section 4 discussed the more general concerns regarding explainabiity, there are some concerns specific to the judicial context. Many make the argument that accuracy, rather than transparency, is a much more important parameter in various fields (such as medicine);[174] but this position will find little support in the legal context because of the principle that all parties to a dispute must be satisfied that the processes that lead to a judicial decision are fair, even administrative ones.[175] This may be extended to investigating officers and other agencies in criminal cases, where explanatory standards may be applied even to human officials.[176]

Understanding the reasoning for a judgment is a key element of procedural due process, and therefore any algorithms used in the judicial process must meet high standards of explainability. Using more transparent but less accurate algorithms where acceptable or avoiding them altogether until algorithms that meet standards of accuracy and transparency are both preferable to using an accurate but inscrutable algorithm in the judicial context.

169 Turner, Jacob. Robot rules: Regulating artificial intelligence.

170 Turner, Jacob. Robot rules: Regulating artificial intelligence.

171 Turner, Jacob. Robot rules: Regulating artificial intelligence.

172 Turner, Jacob. Robot rules: Regulating artificial intelligence. Springer, 2018.

173 A. K. Kraipak & Ors. Etc vs Union Of India, AIR 1970 SC A

174 London, Alex John. "Artificial intelligence and black-box medical decisions: accuracy versus explainability." Hastings Center Report 49, no. 1 (2019): 15-21.

175 Winn, Peter A. "Online court records: Balancing judicial accountability and privacy in an age of electronic information." Wash. L. Rev. 79 (2004): 307. Conley, Amanda, Anupam Datta, Helen Nissenbaum, and Divya Sharma. "Sustaining privacy and open justice in the transition to online court records: A multidisciplinary inquiry." Md. L. Rev. 71 (2011): 772. 5. Morrison, Caren Myers. "Privacy, accountability, and the cooperating defendant: Towards a new role for internet access to court records." Vand. L. Rev. 62 (2009): 919

176 Brennan-Marquez, Kiel. "Plausible cause: Explanatory standards in the age of powerful machines." Vand. L. Rev. 70 (2017): 1249.

These concerns mean that regulations and procedures must be developed to test, certify, and audit algorithms and the data used in support of judicial decisions to ensure that bias is minimised and that their use is safe and fair for a given use case. Algorithms must meet higher standards of fairness and scrutiny for use within the judiciary. Thus, this requirement presents technical challenges. Independent regulation is necessary for the judiciary to benefit from algorithmic support keeping in mind principles of natural justice and constitutional values.

Another concern is the explainability of algorithmic processes. Since the process by which cases are decided is integral to judicial decisions' legitimacy and binding authority, knowing how a decision was reached is much more important in the judicial context than in others. Explainability and bias in algorithmic decision making, particularly in the judicial context, are discussed in more detail later in this section.

LEGISLATION AND COMMON LAW ARE INSUFFICIENT TO PREVENT ALGORITHMS FROM CAUSING HARM

The motivation for recommending that the judiciary creates within it specific and independent regulatory capacity for its use of algorithms derives from the fact that their inherent characteristics are best suited to monitoring, approval, certification, and auditing by an expert body. The legislature is a body of generalists, lacking in expertise.[177] Also, allowing them to formulate rules for the judiciary could hurt judicial independence, and there is potential for external ideological, partisan, and lobbyist influence.[178]

Allowing principles and laws to be set exclusively through judicial precedent and the application of existing laws and rules has other issues. Without any dedicated regulatory regime for algorithmic accountability, multiple overlapping laws and case laws can be applied to algorithmic harm, and many potentially contentious determinations such as the type of liability (eg. strict liability vs vicarious) of the creator will be highly contested, particularly in the event of conflicting decisions between courts of origin and appellate courts.[179]

Cases involving harm due to advanced algorithms may only reach the court after harm takes place, which means that an opportunity to prevent harm has been lost.[180] As observed by the UK House of Commons Science and Technology Committee,[181] resolving conflicts resulting from algorithms' actions can be

177 Scherer, Matthew U. "Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies." Harv. JL & Tech. 29 (2015): 353.

178 Scherer, Matthew U. "Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies

179 Scherer, Matthew U. "Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies

180 Turner, Jacob. Robot rules: Regulating artificial intelligence. Springer, 2018, Matthew U. Scherer. 2015. 'Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies.' Harv. JL & Tech. 29: 353. Turner, Jacob. Robot rules: Regulating artificial intelligence.

181 UK House of Commons. 2016. UK House of Commons Science and Technology Committee Report on Robotics and Artificial Intelligence, Fifth Report of Session 2016–2017. Available at https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pd accessed 1 June 2018.

expensive and time-consuming.

The opacity of machine learning algorithms and the difficulty in determining liability for their actions mean that small, yet severe harms may take place and courts would not be able to respond in a timely fashion.[182] The complexity of algorithms and the difficulty in discerning causality also means that considerable expertise is required to determine the extent of liability, and most judges lack this technical expertise.[183] Regulators must possess this expertise unless and until standards for scrutability and explainability are developed and enforced, and are capable of being used to conduct audits of compliance.

We have established that a significant level of harm and misconduct can go undetected or un-attributed due to the opacity and complexity of advanced algorithms. Cases involving algorithmic harm that reach courts are a small, non-representative sample of all cases where algorithms have harmed a human. Many known patterns in pre-litigation activities, such as negotiations of out-of-court settlements, will further reduce the number of these cases that are resolved in court. As Turner observes, vendors creating or otherwise providing algorithmic tools, products, and services may be willing to settle out of court with victims of harm to avoid damage to their reputation, and the costs of litigation may similarly deter the victims from filing a suit, making settlement a probable outcome.[184] Many types of harm or dispute, and the technological and contextual factors which caused them, may not receive the comprehensive preventative attention that rules and a regulator could provide. Allowing the law on algorithms to evolve through judicial precedent is therefore inadequate to deal with many of the forms of harms or disputes algorithms can inflict.[185]

Regulatory and organisational challenges resulting from the use of algorithms in the judiciary

The ethical challenges of implementing algorithms within the judiciary raise many questions regarding governance and regulation. Many of these concerns emanate from the use of algorithms in general, and their use in public institutions. However, some concerns are specific to the judiciary. A regulatory framework for the development of algorithms for judicial systems must enable the following:

a. upholding ethical principles and due process,

b. developing means of testing compliance with regulations,

c. ensuring that these tests are flexible enough to apply to unforeseeable developments in algorithms' technological capabilities, and

d. addressing any malicious or accidental harm that is caused by the use of algorithms.

Translating ethical principles into design and action is difficult. Therefore more concrete challenges must be addressed for citizens to trust the expansion of the use of algorithms in the judiciary.[186]

Defining the regulatory unit

Attempting to arrive at a non-controversial definition of artificial intelligence, even for the purpose of general or academic use, is

182 Andrew Tutt. 2017. 'An FDA for algorithms.' Admin. L. Rev. 69: 83.

183 Andrew Tutt. 2017. 'An FDA for algorithms.' Admin. L. Rev. 69: 83.

184 Jacob Turner. 2018. Robot rules: Regulating artificial intelligence. Cham, Switzerland: Springer., Matthew U. Scherer. 2015. 'Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies.' Harv. JL & Tech. 29: 353.

185 Jacob Turner. 2018. Robot rules: Regulating artificial intelligence. Cham, Switzerland: Springer., Matthew U. Scherer. 2015. 'Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies.' Harv. JL & Tech. 29: 353.

186 Brent Mittelstadt. 2019. 'Principles alone cannot guarantee ethical AI.' Nature Machine Intelligence: 1-7.

impossible. As mentioned earlier, doing so for the purpose of regulating its judicial application is much more fraught with controversy, given the level of precision and logical consistency required and the dependence of a general definition on a consensus regarding the meaning of 'intelligence'.[187]

Rights and responsibilities

The body would need to perform multiple roles to achieve the regulatory goal of ethical and responsible use of algorithms. First among these would be the formulation of policy, developing model rules and regulations for the use of algorithms in the judiciary and the legal profession.

As mentioned earlier, there are numerous rights and principles that the judicial process is intended to uphold, including fundamental rights, principles of natural justice, and due process. The development and use of algorithms should be consonant with these rights and therefore specific rights and obligations relating to the use of algorithms should be developed. The following are some indicative examples:

1. The right to be informed that inferences derived from algorithmic data processing are being used in any administrative decision[188]
2. The right to an explanation of how an algorithm arrives at a decision or an inference that supports a decision;
    - the right to know what data was used and how it led to the inference or decision, including their relative weightage, and how the inference or decision might have been different if one or more of the parameters in the data were any different; and
    - how the inference will be used to make or support the making of a decision.
3. The right to a human-only review of decisions with algorithmic input.[189]
4. The right to object to being the subject of a decision with algorithmic input.
    - The right against the use of any decision based on an algorithm-generated psychological or cognitive profile or assessment, including as evidence.[190] While the rights listed above would ideally apply to algorithms in a non-judicial context as well, there is a need to list rights specific to the judicial process in the interest of natural justice and due process.
5. The right to challenge the accuracy of algorithmic tools and the inferences they generate when they are relied upon by courts, law enforcement agencies, and investigation agencies. John Lightbourne suggests that "A defendant should have the ability to provide evidence suggesting that the tools used against him or her are flawed–just as he or she would with any other piece of evidence."[191]
    - The right to all information on the

187 Jacob Turner. 2018. Robot rules: Regulating artificial intelligence. Cham, Switzerland: Springer., Matthew U. Scherer. 2015. 'Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies.' Harv. JL & Tech. 29: 353.

188 Toby Walsh. 2016. 'Turing's red flag.' Communications of the ACM 59(7): 34-37.

189 This should not be an extensive right applicable in all situations, but in situations where the algorithms used are sufficiently advanced an opaque, and when the decision is of high enough consequence. For example, there isn't right to human review of decision

190 This is a modified version of the right in Article 4 of GDPR, which uses a more vague definition - "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person". That could possibly be covered by a general right to object to being the subject of a decision with algorithmic input. This modification is intended to focus on the use of profiling to make generalised assertions about a person's character, or to assign a score for an assessment such as likelihood of recidivism. Such scores and profiles are aggregations resulting from what are essentially complex statistical operations. Recalling the distinction between correlation and causation, it is very important to note that an immense body of well-designed and robust empirical research across multiple disciplines, including sociology, psychology, and economics, would be necessary to claim a causal relationship between an algorithm-generated score and something such as recidivism.

191 Lightbourne, John. "Damned lies & criminal sentencing using evidence-based tools." Duke L. & Tech. Rev. 15 (2017): 327. - A.

process by which the tool was selected for use by the judiciary, including information on the performance criteria used and any evaluations conducted. There is precedent in the USA for using the proprietary nature of algorithms used in the process to prevent such disclosures, and revelations from such disclosures have revealed the tool's performance to be well below standards necessary for use in the judicial context.[192]

A similarly indicative and non-exhaustive list of the obligations of the judiciary, the institution itself, and any third-party vendor who may conduct algorithmic processing on their behalf, whether public like the NIC or a private vendor, may consist of the following:

1. The obligation to hold a wider public consultation, to understand the range and prevalence of opinions on the use of algorithms within the judiciary. This must also include a process of spreading awareness on these technologies and their impact, in context. A part of this includes the publication of all documentation of the development of these systems, including mandated impact assessment and compliance with the applicable privacy regime. They must receive and respond to public feedback and queries, within a defined, reasonable period of time.
2. The obligation to inform any person that they are the subject of an algorithmic decision, in observance of the right stated earlier. This also includes providing them with any documentation on the nature of that algorithmic process, covering the data pertaining to that person that is used in this process as well as how the output of the algorithm will play a part in this decision.
3. The obligation to provide a person with an explanation for how the algorithm produced the output that either made or supported a decision impacting them in some way, but particularly as an outcome of a case in which they are a party. This explanation must meet contextual standards of explainability set by the body itself.
4. The obligation to conduct an algorithmic impact assessment to understand the potential harm that could result from an algorithm's use in making or supporting a decision.
5. The obligation to only use algorithms within the regulatory, legal, and constitutional limits, and to implement appropriate technical, organisational, and procedural safeguards to ensure that these limits are not violated.
6. The obligation to provide for a grievance redressal mechanism to redress any harm resulting from algorithmic decision-making in the judiciary, in which the body would be responsible for providing the Supreme Court and High Courts with technical and policy inputs.

Composition - competencies, representation

A body is necessary to determine the changes to be made to legislations, rules, and practices to ensure that algorithms are used safely and responsibly, in consonance with constitutional rights and principles in the judiciary.

Such a body should be created by statute so that it has the authority to draft regulations on the use of algorithm, alternate dispute resolution (ADR) fora, and other institutions in the criminal justice system that participate

192 "Rashida Richardson, Jason M. Schultz, and Vincent M. Southerland. 2019. 'Litigating Algorithms Lightbourne, John. "Damned lies & criminal sentencing using evidence-based tools." Duke L. & Tech. Rev. 15 (2017): 327. - A.

in judicial proceedings.

Accountability to the judiciary

In order to preserve the autonomy of the judiciary, we propose that the regulator be ultimately accountable to the Supreme Court of India, possibly to a relevant committee of the court such as the Artificial Intelligence Committee, which was constituted to identify and implement AI in various use cases.[193]

Judicial representation

The regulator must be led, on a more operational basis, by a person representing the judiciary and familiar with its needs, values and principles, rules, and procedures. This will ensure that the regulator can effectively ensure that the use of algorithms in the judiciary is in accordance with constitutional values, fundamental rights, and ethical principles and to identify opportunities for the use of emergent technologies to improve access to justice.

Technical expertise

One principal motivation for constituting a permanent regulatory body for the judiciary is, as mentioned earlier, is that judges and court staff lack the expertise to oversee the regulation of algorithms. This expertise is essential for framing technical standards, monitoring the ongoing use of algorithmic tools, enabling technically informed and up-to-date policy, and providing technical inputs to the others in the body since their roles will require a degree of familiarity with current and emerging technologies. The inputs of both private sector experts and academicians would be necessary to understand the theoretical and practical dimensions of algorithm-based decision making systems and their capabilities.

193 The Supreme Court of India. 2019. 'Annual Report of the Supreme Court of India'. The Supreme Court of India. Available at https://main.sci.gov.in/pdf/AnnualReports/Supreme_High_Court_AR_English_2018-19.pdf

Lawyers/Bar representatives

Lawyers must be represented because of their familiarity with how algorithmic processes will impact both the fairness of the process from a practitioner's perspective and how it may potentially impact litigants. Given their importance as the first point of contact for citizens who seek access to justice, their perspective is indispensable. Ideally, lawyers from different practice areas and all levels of the judiciary should be represented.

Multiple stakeholder representatives, including civil society groups and NGOs

Civil society organisations and NGOs with a background in judicial research and reform, human rights organisations, criminal justice reform, and digital rights should be consulted to understand how algorithms can affect citizens' rights. As algorithmic systems in the judiciary can potentially have profound social consequences across many spheres, those with the appropriate experience, expertise, and motivation to contribute to policy and regulation in this area should be able to do so. Particular attention must be given in this

regard to the representation of disadvantaged groups, especially since social and economic disadvantages have an impact on access to justice.

Research and Operational staff

The body must have dedicated operational and research staff with appropriate expertise for managing administrative matters and researching various aspects of law, data science, and information technology.

Integration with E-Courts and other initiatives

It is important to note that despite implementing technical and procedural measures to imbue algorithmic systems in the judiciary with fairness, this may fail to prevent the systems from embedding historical biases unless the development of these systems is harmonised with a broader reform programme to increase access to justice equitably.[194] A dedicated body can help coordinate this process. The draft Digital Courts Vision & Roadmap of Phase III of the eCourts Project, which proposes the next phase of digitisation of Indian courts, suggests that digitisation initiatives be overseen by a National Judicial Technology Council (NJTC).[195] This body would be responsible for designing and developing technological solutions for the judiciary, built upon a digital platform. It would develop and set technological standards that ensure that the technology being used is compatible with the constitutional, and legal principles. The proposed NJTC would be well-placed to oversee algorithmic accountability within the judiciary, given its representation of citizens and multiple levels of the judiciary and its technological expertise. It would be able to ensure that algorithms can be incorporated into future judicial systems in a way that maximises the benefits of such technology while avoiding the concerns discussed earlier. In keeping with the federal administrative structure, High Court Computer Committees (HCCC) represent the needs of their respective jurisdictions, and would retain authority over development and implementation of digital infrastructure and software modules in the E-Courts project as per the Phase III Vision document. They would therefore have considerable responsibilities with regard to overseeing the use of algorithmic tools, both for High Courts and for District Courts.

Since High Courts have superintendence over all courts in their jurisdiction as per Article 227 of the Constitution of India, they are not bound to adopt any policy for algorithmic accountability that is set by the NJTC. While their representation in the NJTC should help ensure that they have a role in formulating policy that the NJTC proposes, HCCCs would be able to draft or modify their own policy as their needs require. Sharing of knowledge between HCCCs, and between HCCCs and the NTJC, will make this process much more effective.

Regulatory activities

The task of answering critical ethical, legal, and technical questions will be among the most important activities of the regulatory body, even though these answers cannot be static. Linking algorithmic accountability with the data protection regime will be an integral part of this. The regulatory framework depends on what forms of algorithms apply to a given context, given the answers to these questions. Making and codifying key policy determinations, such as specifying situations, decisions, and roles in which algorithms of a given complexity, or indeed

194 Ben Green. 2018. "Fair'Risk Assessments: A Precarious Approach for Criminal Justice Reform.' In 5th Workshop on fairness, accountability, and transparency in machine learning.

195 https://cdnbbsr.s3waas.gov.in/s388ef51f0bf911e452e8dbb1d807a81ab/uploads/2021/04/2021040344.pdf (accessed on 2021 05 10)

any algorithmic process at all, must never be used, is an essential part of policy formulation for algorithmic accountability in the judiciary. Another related policy choice that the body would be responsible for is determining the requirement for human oversight, based on the use case, its sensitivity, the complexity and opacity of the algorithm, and the potential impact on the subject of a decision.

Standard-setting

Given that policy will necessarily be broad and not technology-specific so that it can continue to be applicable to emerging and new technology, the body will need to serve as a standard-setting body. To begin with, the regulator would need to set the technical standards to enable sophisticated processing of judicial information in the first place. This entails selecting and specifying a 'markup language' that renders judicial documents readable by algorithms by assigning a category to a document's various elements. For example, this is what would enable a lawyer to search for judgments, pleadings, or any other document for all cases filed under a given section of a given statute within a specified time period. Currently, this has been done by private law databases such as Manupatra and Indian Kanoon.[196] The body would also need to specialise the protocols by which various kinds of judicial data are made accessible to algorithmic tools.[197]

As mentioned earlier, it will need to create objective standards for transparency, explainability, and other principles that algorithmic solutions must comply with, given the context that it will be used. To operationalise these standards, the body must set criteria for testing the extent of an algorithmic tool's compliance with relevant policies and ethical standards. Finally, the body should have the authority to certify and deny certification to any algorithmic tool developed for judicial use, administrative use in the judiciary, and the legal profession, based on their performance on these tests. [198]

Selection, certification, and audit of algorithmic systems

The regulatory framework should define the scope of activities in which the use of algorithms is appropriate and justifiable, based on constitutional values, existing laws and practices, and known regulatory difficulties and ethical concerns regarding algorithmic decision-making.[199] Based on these activities, it should define the criteria for the selection of the appropriate algorithm-based tool for the proposed use case and provide a test to ensure that the use is compliant with the selection criteria. The criteria for the quality, completeness, and accuracy of judicial data must be established at the outset, given the significance of data quality in mitigating pre-existing bias. The criteria for evaluating the appropriateness of data for a given use case must be established right at the beginning of development.

Regulating access to the use of judicial data, the use of other data with advanced algorithms by the judiciary, and the use of advanced algorithms in general, must be governed by a judicial privacy framework capable of addressing the risks and vulnerabilities the

196 Vidhi Centre for Legal Policy. 2019. 'Open Courts in the Digital Age : A Prescription for an Open Data Policy.' Vidhi Centre for Legal Policy. https://vidhilegalpolicy.in/wp-content/uploads/2019/11/OpenCourts_digital16dec.pdf

197 Avinash Ambale. '"To the Law Machine" revisited: A Survey & Analysis of Methods and Techniques for Automation in the Legal World'. India Legal. Available at https://www.indialegallive.com/top-news-of-the-day/news/law-machine-revisited/

198 Matthew U. Scherer. 2015. 'Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies.' Harv. JL & Tech. 29: 353; Andrew Tutt. 2017. 'An FDA for algorithms.' Admin. L. Rev. 69: 83.

199 For an ethical matrix to understand the process of selection of algorithmic applications for use by the judiciary, see Fabrice Muhlenbach and Isabelle Sayn. 2019. 'Artificial Intelligence and Law: What Do People Really Want?: Example of a French Multidisciplinary Working Group.' In Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, pp. 224-228.

advancement of algorithms has created. Measures such as data protection impact assessments[200] should be duly incorporated within these frameworks and accompanied by a complementary impact assessment for algorithms, adapted to the context, use case, and the dataset in question.

Algorithmic audits should be conducted within the judiciary to ensure that algorithms perform as intended and are respectful of rights. The performance of algorithms must be benchmarked against humans currently performing the same tasks, such as scheduling hearings, and performance improvement thresholds must be set to ensure that their use in a given application is worth the cost of development. Efficiency is only one parameter in which the use of algorithmic tools can be used to improve performance. Others include the forms of bias and inconsistency that human officials may be prone to and which algorithms can eliminate. Algorithmic audits should address the ethical concerns discussed earlier and compare proposed algorithmic tools with the pre-existing human processes. Even if an algorithm significantly outperforms humans in a given task, it must meet stringent safety, privacy, fairness, predictability, and explainability standards so that its use does not raise ethical questions.

The body would need to select and/or develop an audit framework for using algorithms in the judiciary. Therefore, one of the first tasks would be selecting the system and procedure of auditing algorithms and then specifying what documentation is needed to establish an 'audit trail'.[201] One example is the DEEP-MAX framework, a scorecard for algorithm-based tools based on a combination of ethical and public policy challenges, including fairness, transparency, auditability, diversity, data protection, and equity, for example. Such a scoring system could potentially be developed especially for the judicial applications of algorithms, and factor in appropriateness of the proposed tool for a given use case. [202]

AI Now proposes a model of algorithmic impact assessment, which includes obligations to conduct a self-assessment to identify and publish potential sources of bias, errors, and harm, and set out mitigation strategies, as well as a more generalised explanation of its workings, comprehensible to a layperson.[203] This approach also includes the obligation to allow legal challenges to the system after the publication of documentation and addressing public feedback, and giving access to the system to external researchers, enabling transparent and independent evaluation.[204]

NITI Aayog has proposed a self-assessment guide that includes many of the same obligations but is more closely tied to the model development process.[205] Although only an abridged version is currently available, the broad outlines of both the NITI Aayog and AI Now self-assessment criteria would be applicable in the judicial context once adapted to include principles of natural justice, due

200 Under the Personal Data Protection Bill, 2019, 'Data Protection Impact Assessments' must be undertaken by entities possessing a quantity data above a threshold of volume, sensitivity, and other characteristics, or who intend to process it with 'new technologies', of which advanced algorithms could be one. This assessment must contain the methods used to process data, the method used to do so, the purpose of processing, and the nature of data required. The person or organisation under assessment is also required to submit details of potential harm that would result from their activities, and what safeguards they will implement to prevent it. This assessment is then reviewed by the Data Protection Authority that the Bill proposes to establish.

201 Danielle Keats Citron. 2007. 'Technological due process.' Wash. UL Rev. 85: 1249.

202 Yogesh K. Dwivedi, Laurie Hughes, Elvira Ismagilova, Gert Aarts, Crispin Coombs, Tom Crick, Yanqing Duan et al. 'Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and pol

203 Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker. 2018. Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability. AI Now. https://ainowinstitute.org/aiareport2018.pdf

204 Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker. 2018. ALGORITHMIC IMPACT ASSESSMENTS: A PRACTICAL FRAMEWORK FOR PUBLIC AGENCY ACCOUNTABILITY.

205 NITI AAYOG. 2020. 'WORKING DOCUMENT ON RESPONSIBLE AI FOR ALL'. NITI AAYOG. AVAILABLE AT HTTPS://NITI.GOV.IN/SITES/DEFAULT/FILES/2020-08/RESPONSIBLE_AI_05082020.PDF

process, constitutional values, and concrete means of testing them. Periodical review of impact assessments is an essential part of this obligation.

Grievance Redressal, Ownership, Liability

How the regulatory framework ensures algorithmic accountability will depend heavily on how legislation, rules, and an adjudicatory process, address harm, whether accidental or malicious, that is caused by data breaches, discrimination, bias, or any other violation of human rights and due process resulting from an algorithmic decision. It should also address how to respond to failed audits and discover irregularities or potentially harmful occurrences in an algorithmic system, even when harm occurs. The framework should complement and refer to privacy regulations for judicial data wherever necessary.[206]

The process of engagement with private sector vendors for the development and/ or procurement of these tools, including the liabilities associated with the use of any application, need to be clearly specified. It is foreseeable that the body may recommend adopting private-sector solutions to save on development and deployment costs.

NITI Aayog recommended that strict liability should be avoided in favour of a negligence test and that liability should be limited if a producer of AI tools took necessary precautions during development and implementation.[207] They argue that where multiple parties are responsible for developing a system that causes harm, they should bear the proportional liability and not joint and several liability, and that damages should be apportioned only based on actual harm.[208]

Robert Turner provides an overview of the advantages and disadvantages of different types of liability, recommending that a general regulatory framework should specify the contexts in which each one would apply. For example, vicarious liability has the advantage of enabling the recognition of the limited agency of sophisticated algorithms while still providing for compensation of victims from

206 Lilian Edwards and Michael Veale. 2017. 'Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for.' Duke L. & Tech. Rev. 16: 18.

Marco Almada and Maria Dymitruk. 2020. 'Privacy and Data Protection Constraints to Automated Decision-Making in the Judiciary.' Available at SSRN 3579378.

207 NITI Aayog. 2018. National Strategy On Artificial Intelligence

208 NITI Aayog. 2018. National Strategy On Artificial Intelligence

the producer. However, there is no clarity on the extent of the relationship between the victim and the producer of the algorithm that is needed for the producer to be held liable. Criminal law, on the other hand, may be more closely aligned with the prevailing morality. However, criminal liability requires *mens rea*, meaning that it will be difficult to prove that a human creator is responsible for harm caused by an algorithm as algorithms become more advanced. This creates a 'retribution gap'. Criminal liability is also likely to result in 'overdeterrence' from the development of algorithmic tools.

It is clear that the regulatory framework must address how the various public sector and private builders of algorithmic solutions for the judiciary must be held responsible for their creations, and the degree of the judiciary's own responsibility in each context must also be specified. A grievance redressal mechanism must be provided for, containing means to escalate complaints to the judiciary. Ownership of intellectual property must be factored into the relationships between the judiciary and private vendors of algorithmic tools. In the controversial cases about bias in the COMPAS algorithm, information about and explanations for the working of the algorithm were withheld on the grounds of being a trade secret of Northpointe, the private sector provider of COMPAS.[209] To avoid these situations and to guarantee procedural justice, the terms by which the judiciary can engage private sector developers must specify terms of ownership and disclosure of the algorithmic system itself, to render it transparent when necessary for judicial proceedings.

Public Consultation and Engagement

Engagement with citizens, judges and other judicial stakeholders, lawyers, and experts from academia and the private sector will be critical for these regulatory concerns to be effectively addressed. At the outset, continued engagement with these stakeholders must be planned so that their feedback will shape the policy for use of algorithms in the judiciary. The normative consensus is well recognised as a requirement for effective governance of algorithm-based decision making tools,[210] and regulations must be both understood and supported by citizens for them to be effective.

Training and education within the judiciary

To equip judges to adjudicate algorithmic accountability, the regulatory body can utilise its expertise to develop training programs and educational materials and provide support to judicial academies. It is possible that given the pace of technological development, common law will be important in establishing a framework for algorithmic accountability,[211] and therefore it is all the more important that judges are given adequate support and training in this regard. As the use of algorithmic tools expands within the judiciary, particularly for administrative and clerical purposes, the body could also develop training materials and programmes for the registry and clerical staff where necessary.

Implementation

The regulatory activities discussed above may be split into a loose grouping of stages to serve as a template for the process of overseeing and regulating algorithmic tools. Some concerns must be addressed at the level of design and objectives. Other concerns pertain to the development and implementation which has already taken place (including the development predating

209 Rebecca Wexler. 2018. 'Life, liberty, and trade secrets: Intellectual property in the criminal justice system.' Stan. L. Rev. 70 : 1343.

210 Urs Gasser and Virgilio A.F. Almeida. 2017. 'A Layered Model for AI Governance.' IEEE Internet Computing 21(6) (November): 58–62. doi:10.1109/mic.2017.4180835.

211 Ashley Deeks. 2019. 'The Judicial Demand for Explainable Artificial Intelligence.' Columbia Law Review 119(7): 1829-1850.

any regulatory framework). Some concerns can only be dealt with through continuous monitoring and oversight of algorithmic tools once they are operational. In each stage, regulatory mechanisms must be adapted to deal with these issues, both before and after a regulatory body such as the NJTC is created and the HCCCs gain the capacity to effectively regulate this area.

Before development

The initial concerns which would need to be addressed are at the 'macro level', concern strategy and institutional goals. The development of algorithmic tools and software applications must ideally occur once the judiciary adopts a policy for algorithmic development that addresses the concerns discussed above and demarcates the mission, purposes, and applications for which algorithmic software tools may be developed. The terms of reference of the NJTC/HCCCs if and when they are established, should explicitly demarcate their responsibilities in this area. The strategy and policy for algorithmic development in the judiciary must provide for development based on constitutional, legal, and jurisprudential considerations, as well as principles for evaluating whether development has addressed these. This would formalise rights and obligations,[212] oversight responsibilities of the NJTC/HCCCs and the necessary regulatory capacity,[213] and the essential processes of public consultation and grievance redressal. Policy is necessary to assign individual responsibilities both for oversight and for adverse events resulting from algorithmic processes and determine the means and process of compensating those affected.[214]

Irrespective of the existence of such a policy, any development project must involve a process of adopting technical standards and metrics to assess whether the proposed application has addressed ethical and legal concerns. This should be a transparent and open process involving extensive public consultation and peer review. This is the stage at which the key questions of how ethical principles, procedural law, and any algorithmic accountability policies are translated to code and algorithms must be closely monitored, both internally by the NJTC/HCCCs and externally, as errors in doing so are a significant cause of harm inflicted by algorithmic systems.[215] Metrics should also be adopted to indicate compliance with other policies that would affect the use of algorithmic tools, such as privacy regulations.[216]

The NJTC should ideally create a 'sandbox', a safe environment in which to develop algorithmic tools. In a sandbox, developers operate under close supervision,[217] enabling the NJTC/HCCCs or a vendor appointed by them to identify concerns and vulnerabilities, but also to explore the potential of algorithmic tools.[218] Sandboxes enable third parties to test and train algorithms without needing actual access to the raw data, enabling the development of algorithmic tools to increase access to justice without compromising privacy.[219]

At the time of conception of an algorithmic tool

212 Michael Veale and Irina Brass. 2019 'Administration by algorithm? Public management meets public sector machine learning: Public Management Meets Public Sector Machine Learning,' in Karen Yeung and Martin Lodge (eds.), Algorithmic Regulation, Oxford: Oxford University Press

213 Michael Veale and Irina Brass. 'Administration by algorithm? Public management meets public sector machine learning: Public Management Meets Public Sector Machine Learning,

214 Catrina Denvir, Tristan Fletcher, Jonathan Hay, and Pascoe Pleasence. 2019. 'The Devil in the Detail: Mitigating the Constitutional & Rule of Law Risks Associated with the Use of Artificial Intelligence in the Legal Domain.' Florida State Law Review, (47) 29.

215 Danielle Keats Citron. 2007. "Technological due process." Washington. University Law Review (85) 1249.

216 DAKSH. 2021. 'Paper II - Regulatory Framework for Data Protection and Open Courts'

217 Hilary J. Allen. 2019. 'Regulatory sandboxes.' The George Washington Law Review, 87: 579.

218 Jacob Turner. Robot rules: Regulating artificial intelligence.

219 Hunt, Tyler, Congzheng Song, Reza Shokri, Vitaly Shmatikov, and Emmett Witchel. 2018. 'Chiron: Privacy-preserving machine learning as a service.' arXiv preprint arXiv:1803.05961.

for judicial administration, the first question to address is the necessity for such a tool, and comparing it against alternatives. Since development of automated applications requires time and resources, if the performance of a given system is already satisfactory, those resources may be instead directed towards developing software to address processes that fail to meet performance standards.

If there is a strong case to develop the application, then the next task would be to systematically compare the expected benefits of that system with alternatives, which may or may not be algorithm-based systems. Consider the example of a proposal to introduce a tool to support case management by identifying which legal cases are likely to need more time to dispose of. While different alternatives for algorithmic tools should be compared, other measures, such as passing and enforcing precise, technology-neutral case management rules should be explored, as should options which combine both approaches. As part of this comparison, the nature of human involvement required should also be studied, ranging from a human actor such as a judicial officer merely providing approval or authorisation for a decision reached by an algorithmic process (which we will refer to as "automation") or more actively participating in the decision-making process with algorithmic assistance (which we will refer to as an "augmented" decision-making process).[220]

If the project clears this stage, estimates of resource requirements and budgeting must be prepared and evaluated. This must be accompanied by an assessment of personnel changes, training, and change management is necessary, and if so, to what extent.

During development and implementation

If the NJTC/HCCC finds it feasible to provide these requirements and if, the project should be tested and demonstrated, first in a sandbox, and then in a pilot study. Prior to the sandbox test, the algorithms/model selection and data (including training data and test data) to be used in the pilot should be made public, as should the plan for documentation and reporting of each stage of the pilot. A key part of the pilot should be seeking feedback from litigants, lawyers, witnesses, and other participants periodically.

If the judiciary has not adopted a broad-based policy on algorithms when the development of an algorithmic tool is initiated, the NJTC/HCCC must provide a legally binding guarantee of the rights of litigants, lawyers, and other participants, and affirm their own responsibilities (both rights and responsibilities were discussed earlier). It should also indicate the rights and the responsibilities of the NJTC/HCCC and/or vendor with regard to the judiciary's regulations on privacy.[221]

All documentation, training and test data should be published, and the contents of this documentation should be specified in the regulatory framework. It should include software code, datasets and datasheets which document key attributes of the datasets in a specified format, development and research methodologies, training materials, and any information and research that was relied upon in the development process.[222] This documentation *must* contain an explanation of the workings of algorithms, the nature of insights that will be derived from the data, how they are intended to be used to support administrative decision-making, and

220 Michael Veale and Irina Brass. 'Administration by algorithm? Public management meets public sector machine learning: Public Management Meets Public Sector Machine Learning,

221 Paper II - Regulatory Framework for Data Protection and Open Courts

222 Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2018. 'Datasheets for datasets.' arXiv preprint arXiv:1803.09010

must detail potential sources of bias and all methodologies that are employed to address them.[223]

For systems developed and implemented by a private vendor appointed by the NJTC/HCCC, the contractual relationship between the NJTC/HCCC and the vendor, the procurement/appointment process that was followed, and the nature of the vendor's responsibilities (and liability) should be published. The vendor should maintain copies of all records as described above. Copies should be provided to the NJTC/HCCC, in order for it to be able to provide transparent and informative responses to applications under the Right to Information Act, 2005 (RTI Act). Contracts for engaging any vendor should prevent them from asserting legal claims against any researchers who conduct research on the algorithmic system. This should apply only to research that is conducted in academic interest or by either the State or civil society in the public interest.[224] Any vendor contracts should be updated to be compliant with the regulatory framework.

Since standards and metrics for performance and compliance with the ethical component of the regulatory framework would have been established in the previous stage, data on the pilot study should be frequently published and updated to inform citizens, the NJTC/HCCC, and other stakeholders. Complete documentation and reports should be presented to the NJTC/HCCC and published in an accessible manner for review, and in support of public consultation, to evaluate the effectiveness of the algorithmic tool/system.

Projects which began before the implementation of any regulatory framework should be re-evaluated, and modified accordingly. The use of tools which fail to meet these standards, and which cannot be modified to be compliant, should be discontinued.

After full-scale implementation: ongoing continuous regulatory activities

The NJTC/HCCCs and supporting offices, such as technology offices proposed in the Vision Document, should be endowed with the resources to be conduct research on, and anticipate, emerging regulatory challenges that could render present or earlier systems vulnerable, even if security protocols are adequate for present threats to algorithmic systems. Periodic review, and if necessary, revision of standards will be necessary for a robust and effective regulatory regime.

Any vulnerabilities, data breaches, or any modifications made to the tool after its introduction, should be communicated to all individuals to whom the data, or any insights that the algorithm generates, pertains to. If the tool is developed or operated by a vendor, they must communicate such change to both the people affected as well as the NJTC/HCCC.

The NJTC/HCCC should periodically conduct algorithmic audits to verify that the tool is performing as intended and meets the benchmarks adopted at the outset. The data used to conduct these audits should be made available to enable independent verification of their outcomes. Vendors should be contractually obliged to cooperate with these audits.

In addition to the performance of tools themselves, the NJTC/HCCCs should ensure that mechanisms for grievance redressal and public engagement are working efficiently and that their responses are being fed into improvement of future algorithmic systems.

223 AI Now Institute (2018). Algorithmic Accountability Policy Toolkit. AI Now Institute, available online at https://ainowinstitute.org/aap-toolkit.pdf

224 AI Now Institute (2018). Algorithmic Accountability Policy Toolkit. AI Now Institute, available online at https://ainowinstitute.org/aap-toolkit.pdf

63 Palace Road, Vasanthnagar, Bengaluru 560052

+91 080 4219 0893 | info@dakshindia.org