SAGE Research Methods: Doing Research Online

How to Study Automated Decisions and Algorithmic Injustice in Online Spaces

Contributors: Divij Joshi

Pub. Date: 2022

Product: SAGE Research Methods: Doing Research Online

DOI: https://dx.doi.org/10.4135/9781529608410

Methods: Algorithms, Big data, Training data

Keywords: human rights, law, social media, reverse engineering, political philosophy, transparencies, social science

Disciplines: Anthropology, Business and Management, Criminology and Criminal Justice, Communication and Media Studies, Counseling and Psychotherapy, Economics, Education, Geography, Health, History, Marketing, Nursing, Political Science and International Relations, Psychology, Social Policy and Public Policy, Social Work, Sociology

Access Date: July 26, 2022

Publishing Company: SAGE Publications, Ltd.

City: London

Online ISBN: 9781529608410

How to Study Automated Decisions and Algorithmic Injustice in Online Spaces

Learning Outcomes

By the end of this guide, readers should be able to:

- Understand why it is important to study algorithms and automated decision-making online
- Articulate theories and approaches for understanding algorithmic harm and algorithmic justice based on existing literature
- Practice cross-disciplinary reading and analysis of studies of algorithmic systems and their social consequences
- Identify, design, and develop methodologies for studying algorithmic systems online
- Understand and address the challenges involved in studying algorithmic systems online

Introduction

There is a growing recognition that algorithms and automated decision-making are increasingly prevalent and consequential determinants of culture, politics, economy, and other areas that are of concern to researchers studying the internet and networked societies. In part, this recognition stems from growing concerns about

how algorithmic systems can cause harm to individuals or society and contribute to forms of injustice.

Understanding, researching, and responding to this phenomenon of algorithmic harm and injustice requires a researcher to delve into cross-disciplinary scholarship and research methods which interrogate algorithmic systems as a site of critical research. A researcher must understand how these algorithmic systems in different contexts shape culture and society and what values they dislocate, unsettle, or reproduce.

This brief guide is intended to introduce researchers to fundamental concepts and research methods in the emergent, nascent, and widely interdisciplinary terrain of studying 'algorithms'. It will explain why 'algorithms' are important to consider as an object of social science research and particularly how they disturb widely held values, including conceptions of ethics, fairness, transparency, accountability, and justice, across contexts. It will introduce students to various methodologies for conducting research into algorithmic systems online, with a focus on understanding how these systems might be examined for their impact on various values.

Conceptualising Values and Harm in Algorithmic Systems

To understand and conceptualise algorithmic injustice, we must start with the concept of the algorithm. What is an algorithm? A definition taken from a computer science textbook might describe an algorithm as 'any well-defined computational procedure that takes some value, or set of values, as *input* and produces some value, or set of values, as *output*. An algorithm is thus a sequence of computational steps that transform the input into the output'. (Cormen et al., 2009). Algorithms

(as mathematical processes) are commonly used to sort and classify data or 'optimise' outputs according to the criteria identified by its developers. These functions (and indeed, the very act of translating and formalising human constructs and criterion into computational code) are value-laden exercises (Friedman & Nissenbaum, 1996).

However, social scientists are concerned with algorithms as *social,* or rather, *sociotechnical* objects and systems – implying that algorithms (and their material forms – as in computational software) mutually shape and are shaped by the specific social and cultural contexts within which they are created, utilised, and studied. As Gillespie notes, when we talk about algorithms (as in the 'social media algorithm', or the 'facial recognition' algorithm), we are usually referring to the wider assemblages of the software or source code, the databases that it relies on to operate, the organisational context in which it is created or used, the people responsible for designing multiple aspects of these technical elements, and particularly how each of these entities are interrelated and co-constituted with each other. This requires us to consider algorithms not only as discrete, technological artefacts but as complex, sociotechnical assemblages, which must be studied as such (Gillespie, 2016).

With the increasing prevalence of digitisation, computerisation, and data analytics, these algorithmic assemblages are now commonly used in contexts where their use has consequences for individuals, groups, or society at large. In particular, we are concerned with contexts in which algorithms are used in decision-making to aid or replace the human decision-making capacities and to motivate action taken on the basis of the outputs of an algorithmic process. Contemporary examples

of algorithmic systems in decision-making contexts include content moderation algorithms used in online platforms to present information and media, facial recognition algorithms used in law enforcement, machine-learning algorithms used to identify fraudulent behaviour, or algorithms used to generate credit scores for individuals.

With the increasing relevance of algorithms in contemporary networked societies, a few different frameworks have been proposed to conceptualise and analyse the impacts of algorithmic systems on human values.

Fairness, Transparency, and Accountability

An influential framework for evaluating algorithmic decision-making has been through the lens of specific values of fairness, transparency, and accountability.

Concerns of 'fairness' in algorithmic systems were initially motivated by studies of how biases arise in the design and use of computer systems, such as the design of flight booking web pages, or search engines (Friedman & Nissenbaum, 1996). The use of algorithms to sort information necessarily entails classification, and such classification can produce undesirable kinds of discrimination. Such discrimination can lead to allocative harms – where the access to resources or opportunities may be withheld – as well as representational harm – where the discrimination results in perpetuating and reinforcing harmful cultural perceptions (Crawford & Paglen, 2019). Algorithmic systems might also be perceived as 'unfair' when the outputs of algorithmic systems are arbitrary, inconsistent, or are not performing within the parameters established by their designers.

Several of the concerns that motivate researchers of algorithmic harm arise from Machine Learning (ML) systems, which are a category of algorithmic model which utilises statistical optimisation techniques by learning features and attributes within a set of historical data (training data), and applying the learned functions to future instances of data processing. As we shall see throughout this guide, ML poses particular challenges for researching and conceptualising algorithmic harm and injustice.

With the increasing use of ML systems, designers and researchers of ML systems used in socially consequential areas have begun to examine how algorithms which categorise people for the purpose of credit, employment, etc. might produce or reproduce harmful biases. By learning from historical examples and generalising to future instances of data, ML algorithms can reproduce discrimination and biases in historical data while presenting the results as objectively produced. Such discrimination can occur from the kinds of datasets that an algorithm trains from, the way that the data is organised into 'labels' or 'features' that the algorithm should learn or the kind of statistical operation that the algorithm conducts. Furthermore, as ML algorithms find nonobvious relationships between instances of data that it has trained on, even if clearly discriminatory categories are not used in making classifications, the algorithm can generate discriminatory categories through data that act as proxies for undesirable discrimination (Barocas et al., 2019). In such situations, the considerations and criteria an algorithm uses for classification are considered biased, arbitrary, or irrelevant for the purpose for which it is being utilised.

Notions of fairness in this literature vary widely, emerging from different political, social, and legal concerns (Binns, 2018a) and have to a large extent focussed on how the technical elements of an algorithmic system might be designed to be fairer, such as the employment of technical measures into algorithmic systems for debiasing. These concepts of fairness have been critiqued as failing to recognise broader concerns relating to the fundamental limitations of quantification and data analysis or the broader social contexts or environments within which algorithmic systems might produce discriminatory outcomes, regardless of the technical fairness or accuracy of a system (Gürses et al., 2018; West, 2020). Some scholars have urged for more critical consideration of how discrimination or injustice is produced by 'de-centering technology' (Peña Gangadharan & Niklas, 2019) and being cognizant of how and why algorithmic systems might be used in ways that cause discrimination and harm.

Transparency concerns arise in algorithmic systems for a host of reasons, relating both to the technical nature of algorithmic systems and the broader political, legal, and organisational contexts within which they are embedded. The lack of transparency in the process of algorithmic decision-making has led to their popular conception as 'black boxes' – where the inputs and outputs are observable, but the inner workings seem inscrutable (Pasquale, 2015). Transparency in this context refers to how algorithmic systems and the decisions that they aid or automate can be observed and made knowable to different audiences. It is posited as an instrumental value which can help in achieving accountability or control over algorithmic decision-making systems.

Algorithmic systems obscure this notion of transparency for a number of reasons. First, there are technical perceptions about the limitations of transparency. Particularly in the case of networked algorithms or systems employing complex statistical methods of ML, the functioning of these systems and the logic they employ are incomprehensible for the purpose of human observation (Burrell, 2016). Second, there are concerns about the political, legal, and organisational logics that prevent information about algorithmic systems from being known. These include, for example, trade secret laws that are resorted to by firms that prevent civil society or even governments from scrutinising algorithmic systems that they employ.

While the above section describes some widely noted concerns about how transparency as a norm is affected by algorithmic systems, the notion of transparency in this context has also been problematised. Important scholarship in this area has questioned transparency as a means of establishing accountability over algorithmic systems, focussing on the limitations of transparency (Ananny & Crawford, 2018) as well as questioning how perspectives of transparency are inherently relational and contextual, with multiple meanings across different contexts (Kemper & Kolkman, 2019).

Accountability refers to a relationship between an actor and a forum whereby the actor has an obligation to justify their conduct and can face consequences from a forum which can elicit responses and pass judgement on the actor (Bovens, 2007). Accountability in the context of algorithmic decision-making is concerned with the question of who bears responsibility for the risks or harms that might be produced by algorithmic systems, and how can they be held to account? There

are particular features of algorithmic systems which make the ascription of responsibility and accountability challenging. These include the autonomous nature of many algorithmic systems, where different parts of the system may operate without human oversight, as well as making decisions which are unpredictable and have not been preprogrammed, particularly in the case of ML systems. Owing to this autonomous nature, it is often unclear how responsibility might be attributed to different human actors involved in different parts of their operation (e.g., those responsible for data collection, algorithmic modelling, operations, etc.) (Elish, 2019). Another reason that responsibility might be obscured is because, particularly in the case of networked systems like algorithms that operate on online platforms, algorithmic systems are both complex and dynamic – interacting continuously with real-time streams of data as well as other algorithms and other environmental factors depending on their context. In such cases, it becomes difficult to attribute responsibility to a particular part or operation of the algorithmic system (Yeung, 2018).

The difficulty in attributing responsibility for the failures of, or harms caused by, algorithmic systems often results in the failure of governance institutions or affected persons to meaningfully seek redress for harms caused to them and the continuation of harmful conduct due to lack of accountability by the actors responsible for algorithmic systems.

Apart from being instrumentally important values that facilitate our understanding and responses to perceived harms and injustices in algorithmic systems, transparency and accountability in algorithmic systems implicate fairness-as-process in

decision-making about people, a value that is also recognised in most legal systems as due process or fair decision-making. Fair procedures and due process requires providing individuals with justifications and explanations for actions that affect them and avenues for contesting or appealing decisions perceived as arbitrary, inaccurate, or otherwise unjustified (Citron & Pasquale, 2014).

Ethics, Law, and Rights-Based Frameworks

While the fairness, transparency, and accountability frameworks have provided an important and critical agenda around which much recent scholarship has coalesced, issues of values in algorithmic systems have also been studied from distinct conceptual lenses grounded within disciplines of moral philosophy, law, sociology and political philosophy, among others. The frameworks focussing on ethics, law, and human rights, in particular, provide important normative evaluations of algorithmic systems.

Mittelstadt et al. (2016) have attempted a map of the ethical implications of algorithms, identifying seven distinct ethical dilemmas posed by algorithmic systems from scholarship on the subject. These include (1) the unjustifiability of proceeding on the basis of probabilistic inferences drawn by statistical algorithmic systems, (2) the inscrutability of algorithmic systems leading to opacity, depriving affected people of agency in decision-making (related to the concern of transparency and accountability identified previously), (3) algorithmic design leading to bias, (4) such bias being acted upon to produce discriminatory outcomes (5) personalisation of algorithmic systems leading to limitations on human autonomy, by fore-

closing information or ways of being in online environments (as seen, e.g., in the accusations of lack of political diversity in social media newsfeeds, leading to reduced democratic participation), (6) algorithmic profiling leading to privacy challenges, and (7) difficulties in inscribing moral responsibility to algorithmic systems and their designers, due to complexity in their structure and operation (Mittelstadt et al., 2016). Scholars have proposed many other ethical frameworks to assess algorithmic systems, and normative assessments of ethical values challenged by algorithmic systems is a constantly evolving and much-contested area (see generally Dubber et al., 2020).

Beyond ethical frameworks, algorithmic systems have also been examined from the lens of their implications for legal rights and human rights. While it is beyond the scope of this guide to extensively map out the rights and values within various legal systems that are affected by algorithmic decision-making, a few areas which are particularly relevant to studying algorithmic harm and injustice.

Yeung (2018) have argued for centring human rights concerns in normative evaluations of algorithms and artificial intelligence (AI) systems, arguing that human rights provide a grounded framework for establishing norms around AI, which are also commonly reflected within democratic constitutional orders, and offer a methodological framework within which conflicts between rights can be assessed, offering concrete guidance as to the governance and design of algorithmic systems. Some recent studies have examined how particular human rights might be implicated by algorithmic systems and AI, including, for example, reports by the Council of Europe (Leslie et al., 2021) as well as by UN Special Rapporteur David

Kaye (United Nations OHCHR, 2021) which examine the implications for the right to a fair trial, the right to privacy and data protection, the right to freedom of expression (implicated when algorithmic systems filter and moderate online content), the freedom of assembly (by profiling protestors and foreclosing assembly), and the right to effective remedy (owing to the absence of accountability), among others.

Given that many contemporary algorithmic systems that we are concerned with operate on databases of personal information, much of the legal scholarship engaging with algorithmic harm examines it from the lens of *personal data protection and privacy law*. Barocas and Nissenbaum (2014) have examined how data analytics within algorithmic systems challenge core concepts of privacy law – including the principle of consent and the right to anonymity. Hildebrandt (2019) examines how algorithms challenge fundamental concepts of European privacy law, including fairness and transparency obligations for data processing enshrined in data protection regulation. Some frameworks for examining biased decision-making in algorithmic systems are explicitly grounded within frameworks of *discrimination and equality law*. Barocas and Selbst (2016), for example, look at how algorithmic systems used in employment screening might offend equality treatment obligations under US civil rights and nondiscrimination law. Finally, important scholarship examines how algorithmic decision-making challenges core concepts of *procedural justice* – including procedural requirements for requiring explanations or justifications of certain categories of decisions, as well as the right to secure a remedy from the violation of a right, particularly in situations where such systems

are used by states on their citizens or residents (Citron, 2007).

Section Summary

- Algorithms are technical objects but should also be studied as broader sociotechnical assemblages located in specific cultural and organisational contexts.

  Algorithmic systems are influential in a wide arena of decision-making about people.

- A number of approaches exist for conceptualising the impact of algorithms on human values. These include normative frameworks assessing the fairness, transparency, and accountability of algorithmic systems; frameworks examining their ethical impact; and frameworks examining their impact on legal rights and values, among others.

Methodologies for Studying Algorithmic Harm in Online Spaces

The previous section described some of the implications of algorithmic systems on important human values, including how they lead to harm and failures of justice. These also indicate some of the reasons that algorithmic systems are challenging objects to study. Owing to the lack of transparency, algorithms are often proprietary and closed off to researchers, or otherwise difficult to parse for non-technical researchers. Moreover, even visibility on some features of algorithms –

for example, the variables it takes into consideration – may not present a clear account of what such variables imply (for instance, if a social media feed takes into account political preferences, how are these categories of preference determined?). In complex, dynamic systems, it may be difficult to locate an algorithm as the object of a particular study or attribute causality to a particular part of an algorithmic system. Furthermore, as Seaver (2019) notes, the dynamic and highly personalised nature of algorithms in online systems makes it incredibly difficult to design experimental studies on an unstable object (namely, the online platform which is constantly changing owing to dynamic algorithms Seaver (2019).

A growing body of interdisciplinary literature provides some insight into methodologies for conducting research on algorithms and their consequences, each of which attempts to contend with these challenges.

Reverse Engineering

Diakopoulos (2015) advises us that research into algorithmic systems and their impacts on values is 'best served by a flexible and scrappy' methodological inclination and proposes a theory of reverse engineering to study consequential algorithmic systems. Reverse engineering, here, is 'the process of articulating the specifications of a system through a rigorous examination drawing on domain knowledge, observation, and deduction to unearth a model of how that system works'. (Diakopoulos, 2015) Broadly, reverse engineering is a process whereby a set of inputs into an algorithm is compared with its outputs, in order to generate a mental map or a theory of its behaviour. The method would vary according to the observability of inputs and outputs for the researchers (Trielli & Diakopoulos,

2019).

Examples of ‘reverse engineering’ algorithmic systems to uncover harm include journalistic investigations by ProPublica into a recidivism risk prediction algorithm used in making bail decisions for criminal defendants. In this case, the researchers had access to the algorithms outputs and by comparing output information with other publicly available information, they were able to draw inferences about how the algorithm may have used race (or proxies for race) as a determinant of its recidivism risk score (ProPublica, 2021).

Algorithm Audit

Similar to reverse engineering, an algorithm audit is an experimental procedure that tests algorithms ‘from the outside’, based on information that is publicly available and without the explicit cooperation of the online platforms. Conceptualised by Sandvig et al. (2014), this method builds on the history of social science audits, especially in the United States, which focussed on studying racial disparities in housing and credit opportunities, and examines whether algorithms might exhibit undesirable behaviour. These algorithm audits entail systematically using controlled inputs to observe variation in outputs produced by online algorithmic systems.

In their influential paper on designing algorithm audits, Sandvig et al. (2014) identify a number of ways in which an audit study for algorithms may be designed:

1. Code Audit – researchers who have obtained access to the source code of

an algorithm can probe the software to examine its functioning. Kitchin (2017) also proposes that researchers should examine pseudo-code (the natural language propositions intended to be programmed into an algorithm) as well as source code, by 'sifting through documentation, code and programmer comments and tracing out how the algorithm works to process data and calculate outcomes, and decoding the translation process undertaken to construct the algorithm'.

2. Noninvasive User Survey – this relies on using traditional social science survey methods to identify and probe a set of participants and their interactions with an online platform. This method has the limitation of not clearly being able to clearly link the function of the algorithm with the experiences of participants. Moreover, the design of such surveys requires user sampling that must account for errors in self-reporting by participants, which may be significantly higher when studying sensitive information and personal traits.
3. Direct Scraping – in this method, researchers can directly query an online platform either manually or, more commonly, using an Application Programming Interface (API), which automates queries to online platforms. Here, the researcher can send specific queries to a platform and receive relevant information based on those. Researchers must be wary of the lawfulness of conducting such scraping within their specific jurisdictions, as well of violating the terms of service of online platforms which might prohibit the automated querying of the platform. Ideally, the research should take place within the constraints established by the platform terms of service. Various kinds of APIs are available by different online platforms (Twitter,

Reddit, Facebook) and research tools may be built upon these as well.

4. 'Sock Puppet' investigations and crowdsourced audits – this is an experimental technique involving injecting data to interrogate the platforms functioning, with the aim to manipulate information on the platform to study the effects of the algorithm. A sock puppet investigation involves the use of computer programmes (commonly called bots) to impersonate online users, while the latter involves recruiting testers to probe the online platform or the algorithm.

The design of an algorithm audit is contingent on several factors and subject to important limitations. Audits are generally used to study a particular online platform (such as Facebook or Twitter) or comparisons across a similar service (Google Search or Yahoo Search), identifying algorithmic systems within the context of those services, and are contingent on how information on those services is presented to persons external to it (users, researchers using APIs, etc.). Crucial to the design of an algorithm audit is clearly conceptualising its purpose and the contextual limitations of such a study, which involves mapping the rights and interests of relevant stakeholders and affected persons, the potential ethical concerns which may be raised by an algorithmic system and the kinds of inputs and possible outputs that are most relevant for the scope of the study (Vecchione et al., 2021).

Several important public interest studies of algorithmic systems have employed variations on the audit methodologies described above:

a. Latanya Sweeney's work on online ad delivery systems and Safiya Noble's

analysis of search engine results used controlled inputs to identify racial bias and discrimination based on the algorithmic classification of search results (Noble, 2018; Sweeney, 2013).

b. 'Gender Shades' by Buolamwini and Gebru (2018) analysed disparities in performance of commercial facial recognition algorithms available to use through online APIs, by creating a database labelled for gender and race on which these algorithms operated. The study indicated that popular machine vision algorithms misidentified based on racial and gender characteristics.
c. The 'citizen browser' project by data journalist venture The Markup is using a collaborative audit methodology to understand how and whether social media and video recommendation algorithms filter content and classify individual preferences on the basis of a number of important demographic categories, providing a browser tool to consenting participants to enrol them in a study which assesses demographic disparities in social media and video recommendations (The Markup, 2021). A similar investigation by AlgorithmWatch in Germany relied on crowdsourced outputs of Google search results to show how Google's algorithms filtered the political preferences of users ((AlgorithmWatch, 2021).

Ethnographies of Algorithmic Systems

While audit studies look at specific technical enactments of algorithmic systems in particular online platforms or similar contexts, some scholars have highlighted the importance of studying algorithms ethnographically – as broader sociotechnical systems which are 'culturally enacted' by the human and social practices and

relations through which they are constituted (Seaver, 2017).

Seaver offers some tactics for the ethnographic study of algorithmic systems, including engaging with multiple and heterogenous sites for the field study of algorithms, relying on multisited interviews with subjects, including people designing, engaging with or affected by particular online algorithms, and paying careful attention to the forms of access afforded to the research of algorithms, as well as the positionalities and motivations of ethnographic subjects in their own engagements with the technological object of the algorithm.

Christin suggests enrolling algorithms into the ethnographic study and offers three avenues where such engagement might be generative – algorithmic refraction, namely, examining how particular social contexts and human cultures evolve as algorithmic systems are used; algorithmic comparison, which involves studying the similarities and differences in the perceptions and uses of algorithmic tools in different contexts, to gain insight into lay understandings of the features of algorithmic systems; and triangulation, where online algorithms might be enrolled in examining how online communities evolve, as well as a researcher's own positionality in the ethnographic field, which might be illuminated by how their presence, and difference, is shaped within algorithmically mediated online environments (Christin, 2020).

Important ethnographic work has examined how algorithmic systems are enacted in different social contexts, including those that are relevant to the study of algorithmic harm and injustice online. Brayne and Christin (2021), for example, study

the use of predictive policing algorithms in courts and police departments in the United States, uncovering how these systems contribute to discriminatory or arbitrary outcomes. The thnographic study of Uber drivers and their interactions with the Uber platform by Rosenblat and Stark (2016), including how it shapes power asymmetries between the platform and the drivers, also provides valuable insight into questions of algorithmic harm and justice. Work of Virginia Eubanks (2018) examines algorithmic automation of welfare services in various contexts in the United States, using ethnographic methods such as interviews with case workers using algorithmic systems, people affected by their decisions and organisers or activists protesting such systems, to argue that, examined within the broader economic and political contexts within which they are used, algorithmic systems inevitably profile, punish, and discriminate against people in poverty.

Section Summary

- Studying algorithmic systems is challenging owing to the lack of transparency around their operation, as well as their complex and dynamic nature.
- Multidisciplinary research methodologies have evolved to study algorithmic systems. These include reverse engineering algorithms, algorithm audits that draw on social science auditing methods, as well as ethnographic research into algorithmic systems.

Conclusion

Algorithms are increasingly influential in a range of socially consequential areas, ranging from online media environments to health, finance, and social credit. Contemporary algorithmic systems are pervasive and their influence is only increasing as networked information and computational systems grow. In this guide, we examined how social science researchers might make sense of the growing influence of algorithmic systems from the perspective of their impact on values and, specifically, how they might cause harm and injustice to people.

Studying algorithms and human values requires establishing a conceptual framework within which the operations and impact of an algorithmic system can be deliberated. This guide presented frameworks from interdisciplinary scholarship examining algorithms as both technical and social systems, exhibiting particularities owing to the specific features of computational software, as well as the social contexts in which they are embedded. Some emerging frameworks rooted in technical analyses of algorithmic design examine how algorithms implicate values of fairness, accountability, and transparency, while others take a broader approach rooted in moral philosophy, political philosophy, and law. It is imperative for social science researchers to practice reading across these disciplines in order to understand and construct an analytical framework for algorithmic harm that might apply to their study.

Algorithms are challenging objects and systems to study empirically. This guide presented a few methodologies that researchers can draw on for designing empirical studies of algorithms online while navigating some of these challenges. These frameworks draw from journalistic investigative methods like reverse engineering algorithms to draw inferences about their logics, algorithm audits, which use con-

trolled inputs to study variation in outputs, and tactics and methods for conducting ethnographic studies and analysis of algorithmic systems by situating them in their wider social and cultural contexts.

Multiple Choice Quiz Questions

1. Why is algorithmic harm an important subject of study for social science researchers?

○ **a.** Algorithms are used in multiple computer programmes and scientific developments

Incorrect Answer

Feedback: This is not the correct answer. The correct answer is C.

○ **b.** Social scientists must increasingly use algorithms for computational social science methods

Incorrect Answer

Feedback: This is not the correct answer. The correct answer is C.

○ **c.** Algorithmic systems are increasingly put to socially consequential uses and determine online media environments

Correct Answer

Feedback: Well done, correct answer.

2. What evaluative frameworks exist to study algorithms and harm?

○ **a.** Technical definitions of mathematically computable fairness

Incorrect Answer

Feedback: This is not the correct answer. The correct answer is C.

○ **b.** Legal measures of discrimination and equality, and data protection regulation

Incorrect Answer

Feedback: This is not the correct answer. The correct answer is C.

○ **c.** Multidisciplinary frameworks including computation and information studies, political and moral philosophy and human rights law, among others

Correct Answer

Feedback: Well done, correct answer.

3. Why are algorithms challenging objects for empirical enquiry?

◯ **a.** Social scientists are not able to understand software and code

Incorrect Answer

Feedback: This is not the correct answer. The correct answer is C.

◯ **b.** Algorithmic systems are located in difficult to access spaces which prevent their physical examination

Incorrect Answer

Feedback: This is not the correct answer. The correct answer is C.

○ **c.** Algorithms are complex and dynamic systems, and their material forms in software are obscured owing to technical and organisational constraints

Correct Answer

Feedback: Well done, correct answer.

4. What is an ‘algorithm audit’?

○ **a.** A requirement that algorithms are submitted to a government body to study

Incorrect Answer

Feedback: This is not the correct answer. The correct answer is C.

○ **b.** A method where designers of algorithms are interviewed to understand how a system was built

Incorrect Answer

Feedback: This is not the correct answer. The correct answer is C.

○ **c.** An experimental method where researchers examine code directly or infer the operation of algorithmic systems by providing controlled inputs and observing changes in outputs

Correct Answer

Feedback: Well done, correct answer.

SUBMIT

CLEAR

START OVER

Further Reading

McGregor, L., Murray, D., & Ng, V. (2019). International human rights law as a framework for algorithmic accountability. International & Comparative Law Quarterly, 68(2), 309–343.

Noble, S. U. (2018). Algorithms of oppression. New York University Press.

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown.

Seaver, N. (2019). Captivating algorithms: Recommender systems as traps. Journal of Material Culture, 24(4), 421–436.

Taylor, L. (2017). What is data justice? The case for connecting digital rights and

freedoms globally. Big Data & Society, 4(2), 2053951717736335.

Web Resources

https://fairmlbook.org/

https://facctconference.org/

https://www.ajl.org/

References

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989. 10.1177/1461444816676645

Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness in machine learning. https://fairmlbook.org/pdf/fairmlbook.pdf

AlgorithmWatch. (2021, November). Filterblase geplatzt? Kaum Raum für Personalisierung bei Google-Suchen zur Bundestagswahl 2017. https://algorithmwatch.org/de/filterblase-geplatzt-kaum-raum-fuer-personalisierung-bei-google-suchen-zur-bundestagswahl-2017/

Barocas, S., & Nissenbaum, H. (2014). Big data's end run around procedural privacy protections. *Communications of the ACM*, 57(11), 31–33.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *SSRN Electronic Journal*, 104(3), 671–732. 10.2139/ssrn.2477899

Binns, R. (2018a). What can political philosophy teach us about algorithmic airness?*IEEE Security & Privacy*, 16(3), 73–80. 10.1109/MSP.2018.2701147

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal*, 13(4), 447–468. 10.1111/j.1468-0386.2007.00378.x

Brayne, S., & Christin, A. (2021). Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social Problems*, 68(3), 608–624. 10.1093/socpro/spaa004

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *In Conference on fairness, accountability and transparency* (pp. 77–91). PMLR.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 205395171562251. 10.1177/2053951715622512

Christin, A. (2020). The ethnographer and the algorithm: Beyond the black box. *Theory and Society*, 49(5–6), 897–918. 10.1007/s11186-020-09411-3

Citron, D. K. (2007). Technological due process. *Washington University Law Review*, 85(6), 1249–1314.

Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. *Washington Law Review*, 89, 1.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms*. MIT press.

Crawford, K., & Paglen, T. (2019). *Excavating AI: The politics of images in ma-*

chine learning training sets. AI and Society.

Diakopoulos, N. (2015). Algorithmic accountability. *Digital Journalism*, 3(3), 398–415. 10.1080/21670811.2014.976411

Dubber, M. D., Pasquale, F., & Das, S. (Eds.). (2020). *The Oxford handbook of ethics of AI*. Oxford Handbooks. 10.1093/oxfordhb/9780190067397.001.0001

Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human-robot interaction. *Engaging Science, Technology, and Society*, 5, 40–60. 10.17351/ests2019.260

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, 14(3), 330–347. 10.1145/230538.230561

Gillespie, T. (2016). *Digital keywords: A vocabulary of information society and culture* (Vol. 8). Princeton University Press.

Gürses, S., Overdorf, R., & Balsa, E. (2018). Stirring the pots: Protective optimization technologies.

Hildebrandt, M. (2019). Privacy as protection of the incomputable self: From agnostic to agonistic machine learning. *Theoretical Inquiries in Law*, 20(1), 83–121. 10.1515/til-2019-0004

Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, 22(14), 2081–2096.

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1), 14–29. 10.1080/1369118X.2016.1154087

Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (2014). Frameworks for engagement. In *Privacy, big data, and the public good*. Cambridge University Press. 10.1017/CBO9781107590205

Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., & Briggs, M. (2021). Artificial intelligence, human rights, democracy, and the rule of law: A primer. SSRN Electronic Journal. 10.2139/ssrn.3817999

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 205395171667967. 10.1177/2053951716679679

Noble, S. U. (2018). *Algorithms of oppression*. New York University Press. 10.2307/j.ctt1pwt9w5

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information.

Peña Gangadharan, S., & Niklas, J. (2019). Decentering technology in discourse on discrimination. *Information, Communication & Society*, 22(7), 882–899. 10.1080/1369118X.2019.1593484

ProPublica, M. B. (2021, November26). ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=siiaBuUx_5-LH2f_432kxejIHJI-dlxM

Rosenblat, A., & Stark, L. (2016). Algorithmic labor and information asymmetries: A case study of uber's drivers. *International Journal of Communication*, 10, 27.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. In In Data and discrimination: converting critical concerns into productive inquiry (Vol. 22, pp. 4349–4357).

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2), 205395171773810. 10.1177/2053951717738104

Seaver, N. (2019). Knowing Algorithms.

Sweeney, L. (2013). Discrimination in online ad delivery. *Communications of the ACM*, 56(5), 44–54. 10.1145/2447976.2447990

The Markup. (2021, November26). The Citizen browser project auditing the algorithms of disinformation. https://themarkup.org/citizen-browser

Trielli, D., & Diakopoulos, N. (2019). Search as news curator: The role of Google in shaping attention to news information. In *Proceedings of the 2019 CHI Conference on human factors in computing systems* (pp. 1–15).

United Nations OHCHR. (2021, November26). OHCHR report of the special rapporteur to the general assembly on AI and its impact on freedom of opinion and expression. https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ReportGA73.aspx

Vecchione, B., Levy, K., & Barocas, S. (2021). Algorithmic auditing and social justice: Lessons from the history of audit studies. In *EAAMO '21* (pp. 1–9). 10.1145/3465416.3483294

West, S. M. (2020). Redistribution and rekognition: A feminist critique of algorith-

mic fairness. Catalyst: Feminism, Theory, Technoscience, 6(2).

Yeung, K. (2018). *A study of the implications of advanced digital technologies (Including AI Systems) for the concept of responsibility within A human rights framework* (p. 5). MSI-AUT.

https://dx.doi.org/10.4135/9781529608410