Selected Papers of #AoIR2020:
The 21st Annual Conference of the
Association of Internet Researchers
Virtual Event / -31 October 2020

LEGAL AND ETHICAL PERSPECTIVES ON (BIG) DATA, PLATFORMS, AI AND ALGORITHMS

aline shakti franzke
University Duisburg-Essen

Charles Melvin Ess
University of Oslo

Panel rationale and organization

Since its inception, the Association of Internet Researchers (AoIR) has fostered critical reflection on the ethical and social dimensions of the internet and internet-facilitated communication. These ethical foci are clearly evoked throughout the thematics of the AoIR 2020 conference call, beginning with Power, justice, and inequality in digitally mediated lives; Life, sex, and death vis-à-vis social media; and Political life online.

Concomitantly, Simon Rogerson, Chief Editor of the *Journal of Information, Communication and Ethics in Society* (*JICES*), describes *JICES* as aiming to "…promote thoughtful dialogue regarding the wider social and ethical issues related to the planning, development, implementation and use of new media and information and communication technologies." The Journal thereby offers "necessary interdisciplinary, culturally and geographically diverse works essential to understanding the impacts of the pervasive new media and information and communication technologies."

JICES and AoIR thus share central interests in the ethical and social dimensions of the internet and internet-facilitated communication, and are now collaborating to highlight AoIR conference presentations and papers via publication in *JICES*. As part of this collaboration, we collect here four papers that address these shared interests – with a specific focus on legal and ethical aspects of Big Data. Presuming their acceptance and presentation at AoIR 2020, the papers will be revised especially in light of critical responses received there for inclusion in a special issue of *JICES* devoted to showcasing AoIR ethics work.

Suggested Citation (APA): franzke, a., Ess, C. (2020, October 28-31). *Legal and Ethical Perspectives on (Big) Data, Platforms, AI and Algorithms*. Panel presented at AoIR 2020: The 21th Annual Conference of the Association of Internet Researchers. Virtual Event: AoIR. Retrieved from http://spir.aoir.org.

Paper 1, Towards a Political Theory of Data Justice: A Public Good Perspective, draws on critical data studies and three major political theories of the public good, aiming to synthesize interdisciplinary research on the uses and regulations of digital data in public and political spheres. The authors develop a normative framework outlining the potential public good functions of big data and the necessary normative requirements for the state's rightful collection of large-scale big data, arguing for the state's central role in harnessing large-scale big data to ameliorate digital inequalities and deepen democracy. They offer principles of justice that should guide the regulatory framework of data collection and usage.

Paper 2, Google and Facebook VS Rawls and Lao-Tsu: How Silicon Valley's utilitarianism and Confucianism are bad for Internet ethics, critiques the tech giants' defense of their collection and use of personal data as a questionable *consequentialism* – one that is further entwined with a Confucian-style hierarchical decision-making. Such consequentialism is easily critiqued: predicting the consequences of acts – in this case, of technological development, adoption, and practices such as online data collection – is demonstrably difficult, if not intrinsically impossible. The paper closes by demonstrating a viable alternative to Silicon Valley's utilitarian hegemony through Rawlsian ethics and Taoist rebuttals of Confucianism.

Paper 3, The Jurisprudence of Datafied Law, addresses the growing use of data profiles and algorithmic "decision-making" processes in making legal decisions regarding criminal sentencing, parole, bail, and other jurisprudential outcomes. The rule of law queries the capacity of such systems to adequately address the tension in all democratic systems between *autonomy* and *equity*. The basic assumptions in such systems also evoke basic questions, i.e., whether such profiles are measurements of static, inherent qualities of the individual, or rather a dynamic social metric against which the subject can assess and potentially improve herself. The latter forces further questions as to the proper role of the state in adopting the datafied tools of "surveillance capitalism" to encourage liberal social orders.

Paper 4, A systematic literature Review of ethical Code of Conduct in the field of Internet Research, notes that since 2017, a broad range of documents concerning the ethics of AI, algorithms and big data have proliferated. These documents have mostly focused on basic values, such as autonomy, privacy and transparency. Other approaches focus more on technological processes, such as the moments of data collection, analyses and disemination. This paper provides an extensive literature review and thereby maps the existing landscape of ethical guidelines for these technologies. A total of 90 documents published between the 2017-2020 are analyzed and organized into a taxonomy. The paper addresses three central questions: How are existing guidelines designed? What types of ethical reasoning do they follow? What sorts of ethical schools are represented?

Both individually and collectively, then, these papers directly take up the central interests shared between AoIR and *JICES* in the ethical and social dimensions of the internet and internet-facilitated communication. They offer new insight on legal and ethical aspects of contemporary technologies, some of which will have specific implications for internet research ethics.

Paper 1: TOWARDS A POLITICAL THEORY OF DATA JUSTICE: A PUBLIC GOOD PERSPECTIVE

Chi Kwok
University of Toronto

Ngai Keung Chan
Cornell University

Suggested Citation: Kwok, C., & Chan, N. C. (2020, October). *Towards a political theory of data justice: A public good perspective*. Paper presented at AoIR 2020: The 21th Annual Conference of the Association of Internet Researchers. Virtual Event: AoIR. Retrieved from http://spir.aoir.org.

Introduction

Big data simultaneously enables the state's ability to improve its governance for the improvement of people's living conditions and its ability to abuse its power which threatens the privacy and freedom of democratic citizens. The issue becomes more complicated when we account for the *ownership* and *consent* of big data (Nissenbaum, 2017). It is under this context that boyd and Crawford (2012) posed two central questions about the relations between big data and politics: (1) whether big data can become a public good that is beneficial to people's well-being and good life, and (2) whether the state should be granted the right to collect big data. These questions still lack a *systematic* answer in existing big data literature. With qualification, this paper answers yes to these two questions: Big data can be a public good, but the state can only legitimately use and collect them when it fulfills normative conditions of transparency, fairness, and democratic legitimacy.

The contributions of this paper are threefold: first, it develops *an interdisciplinary political theory of data justice* by connecting three major political theories of the public good (market failure, basic rights, and democratic) (e.g., Kohn, 2020) with empirical studies about the functions of big data (Christin, 2020), thus offering a distinctive perspective explicating why big data should be considered as a public good. Second, it systematically defends the state's right to collect big data from a public good perspective. Third, it offers a normative framework to qualify the conditions under which the state's right to collect big data for beneficial public purposes can be regarded as legitimate. Following Lane et al. (2014), our primary goal is to consider the requirements of justice for "government officials seeking to use big data to serve the public good without harming individual citizens" (p. xi).

Theories of the Public Good and the State's Collection of Big Data

We examine three major approaches of the public good (market failure, basic rights, and democratic) (Kohn, 2020) to explain why big data should be regarded as a public good.

The market failure approach: This approach suggests that when goods are widely beneficial to the public and yet are not profitable, the inability of the market to provide

these goods to a sufficient degree renders the state a legitimate reason to provide them (Kohn, 2020). Consider, for example, real-time traffic data. They could inform drivers to avoid traffic congestion and thereby improving road safety, but it would only be widely used by drivers when the data are freely available to them.

The basic rights approach: Shue (1996) argues that goods related to physical security and basic subsistence—because of their paramount significance to the good life of individual citizens—ought to be provided and guaranteed by the state. Consider the example of pandemic data. The Centers for Disease Control and Prevention in the United States has long been collecting data regarding the spread of various diseases and have used these data to advise people to take appropriate preventive measures. Big data for public health purposes can significantly improve prediction speed and precision (Ginsberg et al., 2009), and hence better protect the lives of many.

The democratic approach: This approach points out that "public goods are goods that provided by the 'public' (e.g., the state) to the 'public' (e.g., citizens or residents)" (Kohn, 2020, p.4) for the sake of deepening democracy. An example is that some local governments in the U.S. collaborate with civic technology firms, such as SeeClickFix and Public Stuff, to provide convenient ways for citizens to share data and concerns over local infrastructural problems directly to local governments. Such initiatives not only strengthen local governments' understanding of communal needs, but also incentivize citizens to actively participate in local governance (Graeff, 2018).

Towards a Political Theory of Data Justice

The paper proposes three central principles of justice in the regulation of the state's *collection* and *uses* of big data.

The principle of transparency and accountability: A central worry about the state's collection of big data is that the processes might infringe on individual citizens' privacy and freedom (e.g., Nissenbaum, 2017). Transparent and open processes open possibilities for public surveillance (e.g., the media and civil society) of the state, thus reducing the chance of the state's abusive use of big data. The monitoring of the state's uses of big data requires an *active* contentious civil society where misbehaviors of the state would be publicly exposed. Thus, the principle requires the state to not only make its own processes of data collection transparent, but also to provide a favorable legal infrastructure for activism against data abuse.

The principle of fairness: The state's collection and uses of big data rely on public finance, and the design of what and how data should be collected is never neutral (Eubanks, 2018). Different designs will result in different social and political groups being benefited. This principle requires the state not only to justify the uses of big data by explaining how it can benefit the public, but also to reasonably explain how the design of data collection does not unfairly skew towards advantaged groups and will not result in negative externalities that harm disadvantaged groups.

The principle of democratic legitimacy: A democratic state's collection and uses of big data can only be legitimate when it is democratically authorized. Given that today's

governments are increasingly reliant on big data for governance (Desrosières, 2002), it is even more urgent to avoid the state becomes a technocracy (Habermas, 2015) in which political problems are deemed the area belongs to political experts who are capable of understanding and harnessing the power of big data. An ability to *see* processes of data collection visible is not equated with an ability to know how they work and should be regulated (Ananny & Crawford, 2018). Therefore, the principle requires not only democratic authorization, but also the massive nurturing of data literacy. A democratic people cannot hold the state accountable to its data abuse and cannot meaningfully authorize the state's collection and uses of big data without understanding what big data is and how it operates.

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.

boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662-679.

Christin, A. (2020). What data can do: A typology of mechanisms. *International Journal of Communication*, *14*, 1115-1134.

Desrosières, A. (2002). *The politics of large numbers: A history of statistical reasoning*. Cambridge, MA: Harvard University Press.

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

Ginsberg, J., Mohebbi, M., Patel, R. *et al.* (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*, 1012-1014.

Graeff, E. (2018). *Evaluating civic technology design for citizen empowerment*. (Unpublished doctoral dissertation). MIT, Cambridge, MA.

Habermas, J. (2015). *The lure of technocracy*. Cambridge, UK: Polity.

Kohn, M. (2020). Public goods and social justice. *Perspectives on Politics*. Advanced online publication. doi: 10.1017/S1537592719004614

Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (Eds.) (2014). *Privacy, big data, and the public good: Frameworks for engagement*. New York, Cambridge University Press.

Nissenbaum, H. (2017). Deregulating collection: Must privacy give way to use regulation? Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3092282

Shue, H. (1996). *Basic rights: Subsistence, affluence, and U.S. foreign policy* (2nd Ed.). Princeton, NJ: Princeton University Press.

Paper 2: GOOGLE AND FACEBOOK VS RAWLS AND LAO-TZU: HOW SILICON VALLEY'S UTILITARIANISM AND CONFUCIANISM ARE BAD FOR INTERNET ETHICS

Morten Bay
University of Southern California

Suggested Citation: Bay, M. (2020, October). *Google and Facebook Vs Rawls and Lao-Tzu: How Silicon Valley's Utilitarianism and Confucianism Are Bad for Internet Ethics*. Paper presented at AoIR 2020: The 21th Annual Conference of the Association of Internet Researchers. Virtual Event: AoIR. Retrieved from http://spir.aoir.org.

Extended abstract

The proposed paper presents an argument in favor of a Rawlsian approach to ethics for Internet technology companies (den Hoven & Rooksby, 2008; Hoffman, 2017). Ethics statements from such companies are analyzed and shown to be utilitarian and teleological in nature, and therefore in opposition to Rawls' theories of justice and fairness. The statements are also shown to have traits in common with Confucian virtue ethics (Ames, 2011; Nylan, 2008).

In contrast to popular perception, American moral philosopher John Rawls did not always denounce consequentialism. He wrote that not taking "consequences into account in judging rightness" would be "irrational, crazy" (Rawls, 1971, p. 30). Rawls' critique of utilitarianism, rather, concerned the *extent* to which utilitarianism relies on consequentialism and also that it is *teleological* (Rawls, 1971).

Hence, viewing the technology ethics and guidelines presented by Internet corporations through a Rawlsian lens raises the question: What is more teleological than companies such as Google, Facebook, and their associated platforms, whose business models entail collecting personal user data and making predictions based on these data? Their stated *telos* is to use the collected data to improve user experiences on services offered to the public for free, and to make contributions to a range of public goods from health care to national security through predictive data analytics. Of course, the data sets are also used to predict the effects of commercial and political advertising, which optimizes the companies' shareholder profits (Zuboff, 2015, 2019).

That the justification for the data collection is presented as the benefits outweighing the harms for the biggest number of people demonstrates the teleological and utilitarian approach taken by these technology companies. The companies' ethics statements are often superficial guidelines with very little adherence to actual ethical practice or theory (Microsoft, 2019; Pichai, 2018). By using what the corporations appear to believe are ethical "buzzwords", these ethics statements often resemble Confucian virtue ethics, in that they present virtues to be adopted without rooting these virtues in empirical

knowledge, ethical theory or presenting a solidly reasoned argument for them (separating them substantially from the virtue-based technology ethics presented by Ess (2011) and Vallor (2016)). Similar to Confucius presenting the ethical necessity of virtues such as order and propriety as somewhat self-evident, the virtues proposed in tech company ethics statements are contextless and theoretically unmoored (Wong, 2012). The ethics practices of technology companies share a characteristic with Confucian virtue ethics in that the companies enforce strictly hierarchical decision-making (Healey & Woods, 2017). The above-mentioned statements and practices are all contingent on the perceived ability of the technology companies to accurately predict the consequences of their actions and the effect of their products. This confidence in predictions coupled with quasi-Confucianist virtue ethics is yet another demonstration of teleological utilitarianism.

Employing an applied ethics method, public ethics statements from Google, Microsoft, and Facebook are analyzed using the work of two opponents of teleological utilitarianism and Confucianism, John Rawls and Lao-Tzu. More than two thousand years apart, Rawls and Lao-Tzu both made compelling and strong arguments against employing conjectures about the consequences of decisions and actions as the foundation for decision-making (Lin et al., 2013; Vuong et al., 2018). Lao-Tzu, likely a pseudonym, did so in the classic Taoist text *Tao Te Ching*, which also contains simple rebuttals of several Confucian virtues. Several arguments emerge from the perspectives of these two philosophers that call the prediction-heavy, teleological and consequentialism-based ethics approach of technology companies into question, including the demonstrable difficulty associated with achieving high accuracy in forecasts of technological development, adoption, and practices such as online data collection (Meade & Islam, 2006).

After showing how the tech industry's utilitarian-Confucian hegemony clashes with Rawlsian ethics and Taoism, these schools of thought are demonstrated as viable alternatives in the construction of technology ethics. The paper argues that these philosophies are particularly viable when considering the ethics of Internet-related technologies, as the communicative, interactive, and participatory nature of the online realm is, arguably, dominated by rapid change.

The speed with which the torrents of changes and transformations flow and thereby constitute the Internet's many domains is not the only thing that makes prediction difficult. As Popper (1945) famously pointed out, a constant increase of human knowledge logically leads to a decreased predictability and a heightened risk of unintended consequences being the outcome. In combination, the speed of change, the production of new information, and the proliferation of the latter, makes the Internet a phenomenon characterized much more by unpredictability than, for example, some examples of hardware development. The paper concludes by arguing how these factors demonstrate that Rawlsian, deontological ethics can be a viable alternative to utilitarianism in technology ethics, perhaps even in combination with elements of Taoist thought.

References

Ames, R. T. (2011). *Confucian role ethics: A vocabulary*. Chinese University Press.
den Hoven, J., & Rooksby, E. (2008). Distributive justice and the value of information: A (broadly) Rawlsian approach. *Information Technology and Moral Philosophy*, *376*.

Ess, C. (2011). Ethical dimensions of new technology/media. *The Handbook of Communication Ethics*, 204–220.

Healey, K., & Woods, R. H. (2017). Processing is not judgment, storage is not memory: A critique of Silicon Valley's moral catechism. *Journal of Media Ethics*, *32*(1), 2–15. https://doi.org/10.1080/23736992.2016.1258990

Henricks, R. G. (2000). *Lao Tzu's Tao Te Ching: A translation of the startling new documents found at Guodian*. Columbia University Press.

Hoffman, A. L. (2017). Beyond distributions and primary goods: Assessing applications of Rawls in information science and technology literature since 1990. *Journal of the Association for Information Science and Technology*, *68*(7), 1601–1618. https://doi.org/10.1002/asi.23747

Lin, L.-H., Ho, Y.-L., & Lin, W.-H. E. (2013). Confucian and Taoist work values: An exploratory study of the Chinese transformational leadership behavior. *Journal of Business Ethics*, *113*(1), 91–103.

Meade, N., & Islam, T. (2006). Modelling and forecasting the diffusion of innovation – A 25-year review. *International Journal of Forecasting*, *22*, 519–545. https://doi.org/10.1016/j.ijforecast.2006.01.005

Microsoft (2019. *Our approach: Microsoft AI principles*. Microsoft. https://www.microsoft.com/en-us/ai/our-approach-to-ai

Nylan, M. (2008). *The five"Confucian" classics*. Yale University Press.

Pichai, S. (2018, June 7). *AI at Google: Our principles*. Google. https://www.blog.google/technology/ai/ai-principles/

Popper, K. (1945). The poverty of historicism, III. *Economica*, *12*(46), 69–89.

Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.

Vuong, Q.-H., Bui, Q.-K., La, V.-P., Vuong, T.-T., Nguyen, V.-H. T., Ho, M.-T., Nguyen, H.-K. T., & Ho, M.-T. (2018). Cultural additivity: Behavioural insights from the interaction of Confucianism, Buddhism and Taoism in folktales. *Palgrave Communications*, *4*(1), 1–15.

Wong, P.-H. (2012). Dao, harmony and personhood: Towards a Confucian ethics of technology. *Philosophy & Technology*, *25*(1), 67–86.

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*(1), 75–89.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.

Paper 3: THE JURISPRUDENCE OF DATAFIED LAW

Dan L. Burk
University of California, Irvine

Suggested Citation: Burk, D. (2020, October). *The Jurisprudence of Datafied Law*. Paper presented at AoIR 2020: The 21th Annual Conference of the Association of Internet Researchers. Virtual Event: AoIR. Retrieved from http://spir.aoir.org.

Introduction

Data profiles drawn from blended private and public digital surveillance are increasingly taking a role in legal decisions regarding criminal sentencing, parole, bail, and other jurisprudential outcomes. (1, 2) Substantive and procedural critiques have been leveled at the of deployment of such datafied law, posing questions regarding accuracy, due process, and oversight. (2, 3) However, surprisingly little has been said regarding the jurisprudential implications of such practices. In particular, there has been little discussion of the fit between values assumed in algorithmic analytics and values embodied by the legal institutions that might employ them. Consequently, in this paper, I begin to map out certain core jurisprudential problems raised by algorithmic profiling within the legal system. In particular, I explore the challenges such algorithmic metrics pose for fundamental liberal values of autonomy and equality.

Algorithmic Accuracy

Legal algorithmic scoring has already been the focus of substantial analytical censure. Much of this negative critique has been couched in the language of accuracy, or in reciprocal language of bias. Commentators on algorithmic legal metrics worry that data profiles will reflect an inaccurate portrait of the subject, either by being incomplete, or by incorporating the social indicia of past prejudices. (4, 5) Such concerns are by no means unfounded; surveillance studies warn us that an assembled data profile is not, and indeed cannot ever be, a fully accurate representation of the subject – no model of real phenomena can by definition ever be as informationally complete the subject of its representation. (6, 7) Neither are such concerns trivial; they are of particular salience where algorithmic training data, analytic data, or algorithmic modelling may reflect or re-inscribe racial, ethnic, or other subordinated minority status. (1, 7)

Jurisprudentially, the accuracy argument may raise concerns grounded in the ethical bases for legal judgment, which are typically couched in terms of dessert or utility. Inaccurate bases for judgment may be unjust from the standpoint of autonomy by invoking undeserved sanctions or creating a legal status that does not fit the character of the individual. Simultaneously, an inaccurate basis for judgment may be inefficient from a utilitarian standpoint as invoking sanctions or creating a legal status unrelated to desired behavioral outcomes.

Interrogating Datafied Judgments

The argument from algorithmic accuracy rests upon the assumption that predictive analytics measure a defined quantity in the universe; that measuring a criminal defendant's risk of recidivism or risk of flight is an exercise equivalent to measuring the defendant's height – either correct or incorrect, perhaps within error bars of a few millimeters. I discuss three jurisprudential consequences that flow from this assumption embedded in the algorithmic metric.

First, because the algorithmic score appears objective, any positive or negative change will likely be attributed to the character or actions of the individual being scored rather than to bias, feedback, or variance in the algorithm. (7) Algorithmic metrics appear to objectively reflect chosen behavior, and so become an ethically significant indicator of individual character. Algorithmic calculation thus precipitates a shift from measurement to judgment, and from actuarial judgment to legal consequence. This suggests that algorithmic legal metrics, whether or not intended to be dispositive in a legal decision, will be accorded greater weight than they merit.

Second, the generation of algorithmically tailored legal standards highlights a latent tension that exists in all democratic systems, between the values of autonomy and equity: each individual is to be valued for his or her own distinctive personhood, but at the same time all are to be treated equally before the law. (8) Democratic regulation becomes illegitimate when it is arbitrary, but may be equally illegitimate when impersonal and calculated – too much personal variation violates democratic principles of equality, but too little violates liberal principles of autonomy. (9) An "unjust" legal regime may mean a regime that undervalues either the former or the latter. Legal standards accommodate some of each value, but

Finally, I suggest that algorithmic metrics raise questions regarding legal governance grounded in determinism or in autonomy. Different strategies are implicated depending upon the jurisprudential model of the subject, and whether the data profile is believed to measure a quality of the subject that should be regarded as static or dynamic. Regarding the algorithmic determination as a measure of a static, inherent quality supports a theory of utility or of dessert. Deontologically, the subject may be said to deserve whatever state of character the algorithm indicates; consequentially, remedies directed to static characteristics may be most efficient at encouraging or deterring behaviors.

But if such traits are considered dynamic, open to self-inspection and alteration, as opposed to external punishment or reward, then the algorithmic score offers a mechanism of assessment and a goal for sanctioned improvement. This is of course the type of manipulation that Yeung has cautioned against as "hypernudging." (10) And although behavioral rehabilitation is a recognized value of some types of adjudication, this use of algorithmic metrics raises a separate set of questions as to whether the of "surveillance capitalism" are legitimate tools of state authority. (11)

Conclusion

By illuminating the discontinuities between the values assumed by predictive analytics and those assumed in liberal legal institutions, this paper makes several novel contributions to the literature on law and algorithmic governance. First, it draws together several disparate strands of literature on the social and normative implications of algorithms to offer a framework for assessment of such technologies in the context of legal adjudication. Second, it shows that the current debates over the accuracy of predictive legal metrics overlook core issues surrounding their use in the legal system, suggesting the inadequacy of conventional liberal curatives such as algorithmic transparency or due process. Finally, the paper elucidates the treatment of algorithmically determined profiles as either dynamic or static personal characteristics, and shows how such models implicate the current discussion over algorithmic manipulation that have been primarily directed toward private rather than public actors. These findings point the way to a more complete and fruitful discussion regarding propriety of deploying of algorithmic legal metrics.

References

(1) Margaret Hu, *Algorithmic Jim Crow*, 86 FORDHAM L. REV. 633 (2017).

(2) Sonia Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. Rev. 54 (2019).

(3) Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1256 (2008).

(4) Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671 (2016).

(5) Danielle K. Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014).

(6) Kevin D. Haggerty & Richard V. Ericson, *The Surveillant Assemblage*, 51 BRIT. J. SOC. 605 (2000).

(7) Marion Fourcade & Kieran Healy, *Seeing Like a Market*, 15 SOCIO-ECON. REV. 9, 24 (2017)

(8) Wendy Brown, *Wounded Attachments*, 21 POL. THEORY 390 (1993).

(9) Hans Christoph Grigoleit & Philip Maximillian Bender*, The Law Between Generality and Particularity – Potentials and Limits of Personalized Law* in DATA ECONOMY AND ALGORITHMIC REGULATION: A HANDBOOK ON PERSONALIZED LAW (Christoph Busch & Alberto De Franceschi, eds., forthcoming 2020).

(10) Karen Yeung. *'Hypernudge': Big Data as a Mode of Regulation by Design*, 20 INFO. COMM. & SOC'Y, 118 (2017).

(11) SHOSHANA ZUBOFF, THE *AGE OF SURVEILLANCE CAPITALISM*: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019).

Paper 4: A SYSTEMATIC LITERATURE REVIEW OF ETHICAL CODES OF CONDUCT IN THE FIELD OF INTERNET RESEARCH

aline shakti franzke
University Duisburg-Essen

Suggested Citation: franzke, a.s. (2020, October). *A Systematic Literature Review of Ethics Codes of Conduct in the Field of Internet Research*. Paper presented at AoIR 2020: The 21th Annual Conference of the Association of Internet Researchers. Virtual Event: AoIR. Retrieved from http://spir.aoir.org.

Purpose – As Big Data and Artificial Intelligence/Machine Learning proliferate, calls for ethical reflection emerge. Ethics guidelines play a central role in this respect. While quantitative research on the ethics guidelines of AI has been undertaken by Jobin et al., systematic qualitative analysis of the ethics guidelines pertaining to AI/Big Data has remained lacking.

Design/methodology/approach – Aiming to bridge this research gap, this paper analyses 85 international ethics guideline documents from academia, NGOs and corporate backgrounds, published between the year 2017-2020.

Findings – For meaningful ethics guidelines it is necessary to define underlying ethics approaches, to explicate values and possible harms. Therefore, Virtue Ethical approaches are recommended.

Originality/value – The paper provides fine-grained qualitative insights into the architecture of AI guidelines, which may prove beneficial for developers, academics and regulators.

Keywords – Artificial Intelligence, AI ethics guidelines, Virtue Ethics

Paper type – Extended Abstract

Introduction

The question of how to engage with ethical principles in the design, implementation and usage of AI is not only frequently discussed across the media and by policymakers, but is also at the centre of a mushrooming debate in academia. Robust quantitative research has shown that ethics guidelines for AI prioritise the values of Transparency, Justice and Fairness, Non-maleficence, Responsibility, Beneficence, Freedom and Autonomy, Trust, Sustainability and Dignity (Jobin et al. 2019, p. 395).

Some have raised concerns that these principles are insufficient for ensuring ethical AI (Mittelstadt 2019). The broad field of stakeholders and developed tools make it difficult to oversee the principles and categories involved (Morley et al. 2019). The entanglement between ethics and business has been highlighted (Hagendorff 2019, p.107). It has also been pointed out that both technical implications are frequently missing in ethics guidelines and their legal enforcement has been lacking (Hagendorff 2019, p.111). Most recently, the normative requirements that arise from these ethics guidelines have been highlighted (Stahl, Ryan, 2020).

An extensive qualitative analysis of existing ethics guidelines of AI, however, has been missing. Such an analysis is not only of interest to the academic debate but also to regulators, who seek to determine the soft influence ethics guidelines might or might not have. The guiding interest of this paper is: what can be observed when one examines the extant ethics guidelines of AI in a qualitative manner?

Method

Algorithm Watch, crowdsourced a Global Inventory of AI ethics Guidelines (https://inventory.algorithmwatch.org/). This list was used as a starting point. 167 articles from the years 2017-2020 have been extracted. After studying title, abstracts and the entire guidelines, a total amount of 85 was finally included in the qualitative analysis. Specific values and observations regarding the architecture of the document have been noted for each entry.

Findings

The findings can be grouped into three categories:

1) **Theoretical**

    a) What is meant by ethics in the sample?

    Guidelines frequently use the term ethics without further clarification about what school of ethics informs this usage or what is understood by it. Differing examples, however, can be found (e.g.: https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf)

b) What is meant by specific values in the sample?

Values are introduced without specifying how to define them, creating something of a “value monoculture”. While privacy is frequently mentioned, only a handful of guidelines address de facto privacy breaches and the implications of these in the lives of those affected (e.g. https://ec.europa.eu/futurium/en/ai-alliance-consultation, https://www.montrealdeclaration-responsibleai.com/the-declaration)

c) Utilitarian dominance

If there is any indication of a moral tradition, it is almost invariably a utilitarian one (f.e: https://www.accenture.com/gb-en/company-responsible-ai-robotics)

d) “Dialogical” absence

In most of the approaches dialogical, deliberative elements are missing. Deliberative group focused processes, however, would help to dismantle open questions about how to put ethics into action (e.g. https://dataschool.nl/de/deda/)

2) Individual level

e) What kind of specific harm is imagined in the sample?

If guidelines speak about harm, frequently they don’t specify what kind of harm could actually occur, or what the specific dangers of unethical AI would be (exceptions are: https://fpf.org/wp-content/uploads/2017/12/FPF-Automated-Decision-Making-Harms-and-Mitigation-Charts.pdf).

f) Who might be excluded?

If datasets are referred to, there is often no clear indication about what kind of groups of people might be excluded (exception: https://i.unu.edu/media/cs.unu.edu/page/4453/UNU-MACAU_Data_Marginalization_Flyer.pdf).

3) Institutional and Regulative Level

g) How to follow-up new findings and implement ethical reflection into institutional settings?

Frequently there is no indication of how to operationalise the findings of the ethics reflection in follow-up steps (exception: https://theodi.org/wp-content/uploads/2019/07/ODI-Data-Ethics-Canvas-2019-05.pdf).

h) How to regulate the use of Big Data analysis?

There is on the whole no indication of how to contact legal entities and report on possible findings or discovered “to do’s” within ethics reflection (exception: https://ec.europa.eu/futurium/en/ai-alliance-consultation).

Discussion

These qualitative findings provide a contribution to the discussion on the meaningful design and use of Ethics Guidelines of AI.

Without a clear understanding of the underlying approach to ethics, and the definition of values, ethics guidelines risk to legitimize common practice in the respective field of application, to conceal political positions, to function solely as a public relations strategy and to be without consequences. For example, speaking about privacy alone without specifying what we exactly is meant by it is not sufficient. Undoubtedly, the current difficulty lies in the details of how values could be operationalized. However, even though this point is important, it oversees the necessity for ethical dialogue and deliberation. Ethics guidelines at their best serve to stabilize and distil standards, point toward opaqueness and establish best practices. In order to establish these, more virtue ethical approaches to ethics guidelines of AI, which are sensitive to the specific context could be helpful because they entail more deliberative understandings of what it means to do ethics.

References

Accenture (2020). Responsible AI and Robotics: An Ethical Framework. Last accessed: 28.8.202. Online: https://www.accenture.com/gb-en/company-responsible-ai-robotics

Algorithm Watch (2019). AI Ethics Guidelines: A global Inventory. Last accessed: 28.8.2020. Online: https://inventory.algorithmwatch.org/

Data Protection Authority (CNIL) (2017). How can humans keep the upper hand?: The ethical matters raised by algorithms and artificial intelligence. Last accessed 28.8.2020. Online: https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf

EU High-Level Expert Group on Artificial Intelligence (8.4.2018). Ethics Guidelines for trustworthy AI. Last accessed 29.8.2020. Online: https://ec.europa.eu/futurium/en/ai-alliance-consultation

Franzke, A., Muis,I., Schäfer,T (2017). Data Ethics Decsions Aid. Last accessed 28.8.2020. Online: https://dataschool.nl/de/deda/

Future of Privacy Forum (2017). Unfairness by Algorithm: Destilling the harms of Automated decision making. Last accessed 29.8.2020. Online: https://fpf.org/wp-content/uploads/2017/12/FPF-Automated-Decision-Making-Harms-and-Mitigation-Charts.pdf

Hagendorff, T. (2020). The ethics of Ai ethics: An evaluation of guidelines. *Minds and Machines*, 1-22.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1-7.

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2019). From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, 1-28.

ODI (2019). Data Ethics Canvas. Last accessed: 28.8.2020. Online: https://theodi.org/wp-content/uploads/2019/07/ODI-Data-Ethics-Canvas-2019-05.pdf

Ryan, M., & Stahl, B. C. (2020). Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *Journal of Information, Communication and Ethics in Society*.

Thinyane,M & Christine, D. (2019). A typological framework for Data Marginalization: Identifying forms of marginalization and exclusion in data-intensive ecosystems. United Nations University Institute in Macau. Last accessed: 28.8.2020. Online: https://i.unu.edu/media/cs.unu.edu/page/4453/UNU-MACAU_Data_Marginalization_Flyer.pdf

Universite de Montreal (2017). Declaration for a Responsible Development of AI. Last accessed 29.8.2020. Online: https://www.montrealdeclaration-responsibleai.com/the-declaration