Please cite the Published Version

Compounding barriers to fairness in the digital technology ecosystem

Sandra I. Woolley
School of Computing and Mathematics
Keele University
Staffordshire, UK
orcid.org/0000-0002-7623-2866

Tim Collins
Department of Engineering
Manchester Metropolitan University
Manchester, UK
orcid.org/0000-0003-2841-1947

Peter Andras
Schools of Computing and Eng. & the
Built Environment, Napier University
Edinburgh, UK
orcid.org/0000-0002-9321-3296

Allison Gardner
School of Computing and Mathematics
Keele University
Staffordshire, UK
orcid.org/0000-0002-8483-121X

Marco Ortolani
School of Computing and Mathematics
Keele University
Staffordshire, UK
orcid.org/0000-0001-6759-7698

Jeremy Pitt
Dept. Electrical and Electronic Eng.
Imperial College London
London, UK
j.pitt@imperial.ac.uk

Abstract— **A growing sense of unfairness permeates our quasi-digital society. Despite drivers supporting and motivating ethical practice in the digital technology ecosystem, there are compounding barriers to fairness that, at every level, impact technology innovation, delivery and access. Amongst these are barriers and omissions at the earliest stages of technology intentionality and design; systemic inadequacies in sensing systems that deteriorate performance for individuals based on ethnicity, age and physicality; system design, co-requisite and interface decisions that limit access; biases and inequities in datasets and algorithms; and limiting factors in system function and security. Additionally, there are concerns about unethical and illegal practices amongst digital technology providers: for example, in planned obsolescence and anti-competitive behaviors, failings in data practices and security, and in responses to problematic use and behaviors. It is critical that these failings are identified and addressed to better evolve a fairer future digital technology ecosystem. This paper contributes a perspective on technological stewardship and innovation; it identifies the compounding nature of barriers to fairness in the current digital technology ecosystem, and contrasts these with the non-compounding fairness drivers that, in general, establish minimum requirements.**

Keywords—digital technology, fairness, fairness metrics

I. INTRODUCTION

Our lives increasingly rely on rapidly changing digital technologies but, at every level of technology evolution and delivery, there are barriers and failings that compound to limit the technological equity experienced by substantial populations of individuals. Yet the disadvantages and impacts of digital exclusion in future digital living are considerable, and particularly so when we imagine the expansive nature of future technologies that encompass not only digital devices but also digital assistants, bots and robots; virtual, augmented, and mixed realities; smart homes, smart energy and smart transport; wearables and the wider Internet of Things (IoT); and the interfaces connecting individuals and communities, and data and services.

The importance of ethical practice in society, business and professional life is manifested by the many drivers that encourage and enforce fair and appropriate conduct as, for example, illustrated in Figure 1 by the drivers acting in support of fair technology. Like many drivers of positive behavior, these drivers are rooted in reactive initiatives that have responded to historical failings by establishing minimum requirements and minimum standards. The challenge now, in part, is to achieve responses that, can keep apace with the rapid nature of technological change and, ideally, prevent the permeation of unfair technologies and practices. A further challenge is the inherent insufficiency of 'minimum requirements drivers' to compensate for layers of multi-component barriers that, in essence, act as monotonic decreasing functions to fairness. In the next sections we consider facets of technological fairness before returning to the mismatch between drivers and barriers.
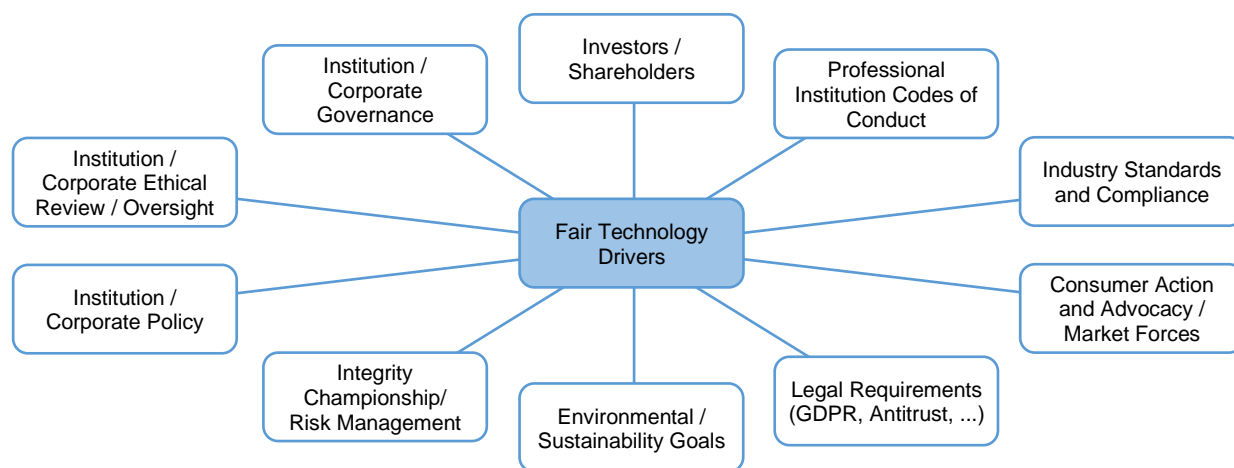

Fig. 1. Fair Technology Drivers

II. FAIRNESS

In this section, starting with contextual fairness, we consider some of the key facets of fairness that underpin our consideration of compounding technological barriers.

A. Contextual Fairness

Notions of fairness are complex and nuanced, and framed by societal and cultural norms. What is considered as fair, just or equitable, therefore, depends on the context and on the dimensions of moral or ethical frameworks, whether they be teleological (goal-based) or deontological (duty-based) as, for example, ideals aligned with Rawls' theorization of justice as fairness [1], where social primary goods are equally distributed unless an unequal distribution is to the advantage of the least favored. But how can we translate and orientate different interpretations of fairness into digital living and the diverse stakeholders, including both users and non-users, affected by the digital technology ecosystem? How can we avoid the omissions, implicit assumptions and unconscious biases that create barriers according to age, ethnicity, gender, ability or physicality that have consequences for technology to be used equally by all members of society?

B. Technological Fairness and Fairness Metrics

Computing and communicating innovations have, for many years, relied on technological 'fairness metrics' to resolve resource allocation problems [2]-[4]. For example, backoff algorithms and traffic prioritization schema to resolve the challenges of fair and appropriate bandwidth and resource sharing to enable local area networking and the Internet communications infrastructure. Such fairness metrics now pervade digital technology devices and infrastructure, in the algorithms and protocols of computing and communicating devices, in the widespread use of throttling and load-balancing, and in the systemic prioritizations of traffic, services and resources. But these low-level, quantitative fairness metrics need to be complemented by higher-level assessments of technological fairness that are both subjective and collective, such as those found in systems of interactional justice [5].

C. Transparency

Transparency is fundamentally linked to fairness and important where there may be concerns about inclusion, accuracy, privacy or trust. In research, transparency is essential for repeatability, and it is increasingly important as researchers intensify use of primary and secondary human data, and share, collate and reuse datasets [6] that ultimately feed machine learning algorithms that themselves can inform decisions of consequence [7]. Of course, transparency is also important across the spectrum of safety-critical systems from implantable medical devices [8] to self-driving cars [9].

Where a lack of transparency combines with misplaced trust in complex systems, there can be disastrous consequences as, demonstrated by the Boeing MAX 737 airplane design and software failings that resulted in passenger airline crashes in 2018 and 2019 [10], [11]. But calamitous technology consequences are not limited to safety-critical scenarios as, for example, demonstrated by the fundamentally flawed UK Post Office accounting software [12] that resulted in *"the UK's most widespread miscarriage of justice"* [13], [14].

There is also recognition of a need for more transparency in IoT systems. For example, there is evidence from smart home technology research that users are frustrated by the inconsistency and lack of transparency about their IoT systems [15]. Though, of course, there is a difficult balance between enabling transparency about data transactions and device status, and information overload. At the same time, transparency is challenging because of the complicated nature of digital systems whose software encompasses apps, device firmware and operating systems (as well as firmware embedded in sensing modules) and the cloud services of data controllers and processors.

D. Interaction in Socio-technical Systems

The enablement of interaction can go beyond transparency, but it is possible to identify at least four layers of interaction in relation to the management of resources in socio-technical systems [16]:

- *delegated* – humans out of the loop: operation is performed by 'smart' automation, which is performed without (or with limited) user intervention, typically being the domain of cyber-physical or multi-agent systems;

- *programmable* – humans on the loop: operational parameters are specified by users, 'smart' automation resolves the constraints, and presents plans to users, for example route finding and load balancing in intelligent transportation systems;

- *interactive* – humans in the loop: automation indicates active intervention is required by humans, e.g., overload prevention by SmartMeters in an islanded community energy system initiating collective action (but the users choose what to turn off);

- *attentive* – humans are the loop: a human participant may occupy multiple different stakeholder roles, especially as an inhabitant of a SmartCity, being called upon to participate in local decisions through citizen assemblies, for example.

III. COMPOUNDING BARRIERS TO FAIRNESS

The compounding barriers to fairness in the digital technology ecosystem are illustrated in Figure 2. This multi-level (A-H) process-oriented model evolved in part from use case examples (and 'non-use' case examples) of unmet needs, failings, and discriminatory and exclusionary practice. For example, use cases considering why all drivers and route-planners don't benefit from real-time traffic data, why all children don't have computers, and why so few elders use and benefit from wearable health trackers?

A. Technology Intentionality and Design

Barriers to fairness that include purposeful exploitation and digital dependence [17] are significant at the earliest stages of the digital technology ecosystem where the intentionality and design of technologies evolve. Omissions at this level also introduce barriers to fairness. For example, the pervasive lack of Equality, Diversity and Inclusion (EDI) across technology research, design, development and financing [18] as well as in participant test cohorts [19] that can fundamentally limit the intentionality, conception and early steer of technology designs and visions and the perceived users and stakeholders.
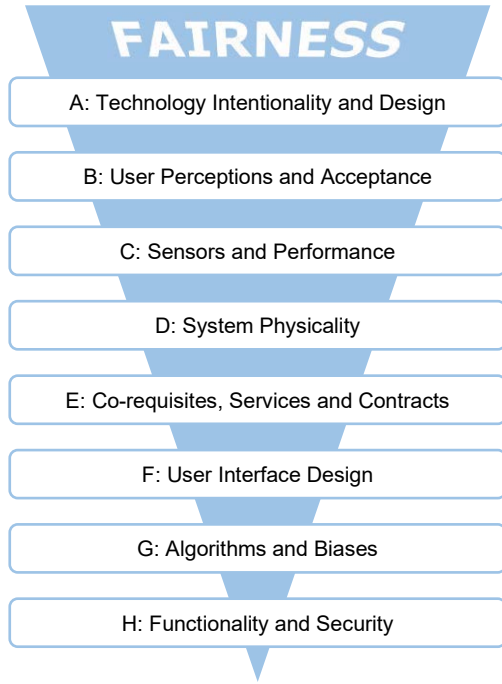
Fig. 2. Compounding Fairness Barriers in the Technology Ecosystem

B. User Perceptions and Acceptance

Where technologies *are* designed for inclusive use, there are user perceptions around ease-of-use and usefulness that are well established as limiting factors for technology adoption and acceptance [20]. These perceptions may stem directly from component factors such as experience or lack of experience, or be influenced by social or cultural factors.

C. Sensors and Performance

IoT systems and other technologies that rely on sensed data can lack accuracy and can exclude individuals based on ethnicity, age and physicality. For example, wearable health trackers count steps less accurately for slow walkers, and optical heart rate sensing performance deteriorates for individuals who are older, have higher BMIs or darker skin tones [21], [22]. But despite inaccuracies [23], [24], inclusivity failings [25] and a lack of evaluation reporting in the academic literature [26], the devices do achieve positive and worthwhile health outcomes and insights for many users. However, it is concerning that these benefits, together with the ability to acquire real-world sense data, have incentivized applications beyond individual well-being into healthcare, research practice, health insurance [27] and corporate wellness [28]. But sensor-related fairness failings are not limited to wearable fitness trackers. Individuals with darker skin tones may experience larger errors when using inappropriately calibrated pulse oximeters and may be unable to obtain soap or water from taps and dispensers fitted with reflective infrared proximity sensors [29], similarly older individuals may not be able to make use of fingerprint sensors [30].

D. System Physicality

The physicality and inflexibility of digital systems can present many barriers to accessible use. For example, wearable device screens and many smartphone and mobile device screens are too small for many elders. Additionally, wearables like other IoT systems are often tethered to smartphones and not accessible via larger desktop screens.

E. Co-requisites, Services and Contracts

Technology dependencies and co-requisites present additional barriers to technology access and adoption. Beyond the immediate technology cost there can be additional subscription costs or co-requisite services and contracts, e.g., a capable smartphone and contract or broadband connectivity Similarly, smartwatches and activity monitors, IoT devices, and smart energy meters often rely on accompanying smartphone apps for their user interfaces.

F. User Interface Design

Beyond the physical constraints of system interfaces such as screen size, there are myriad implicit shallow and deep design factors that influence technology access, usability and adoption. For example, reliance on gesture or interface familiarity [31]. Unfortunately, although usability issues are often recognized and addressed by manufacturers or providers, accessibility and inclusivity are frequently neglected until relatively late in the design process [32].

G. Algorithms and Biases

Algorithmic bias in digital technologies can lead to inequitable outcomes for individuals and demographic cohorts and, result in disadvantage or harm [33], [34]. Algorithmic bias is usually attributed to the inappropriateness or unrepresentative nature of artificial intelligence (AI) training datasets that replicate past and present societal biases [35]. Whilst greater transparency regarding the technology itself is seen as a potential solution to lack of trust it can, paradoxically, further reduce user trust and confidence in the systems [36].

One of the most striking implications of living in a digital society is that while data is constantly generated *about* we individuals from sensors, cameras, wearable devices and social networks, the same data is used to trigger decisions *for us*, with potentially far-reaching consequences [37]. For example, information about spending habits is used to predict solvency, and may directly influence future access to credit. The introduction of algorithmic models for credit scoring, ideally free from the arbitrary judgement of human operators, has often been promoted as the triumph of fairness and objectivity. However, such models have raised concerns, for example, in terms of potential racial bias [38], and it has been shown that businesses founded by disadvantaged minorities have much lower average business credit scores. Similarly, widely used commercial systems for automatic face recognition are prone to discrimination based on race and gender [39]. These findings are doubly worrying because machine learning models carry an aura of objectivity that may render the decisions taken as a consequence of their outcome difficult to challenge, especially for the layperson. On the other hand, when bias becomes apparent, users may lose their trust in the autonomous systems and reject them altogether, losing also all the potential benefits [40]. Indeed, a lack of public trust in the fairness of algorithmic and data-driven systems has increased following a number of prominent examples of systems that have been discriminatory [41]. This lack of trust is compounded by fears that key decision-makers may default to the use of algorithms, (automation bias). Such fears have some foundation, for example users can indeed default to algorithmic advice over peer-advice (unless any inaccurate advice is clearly labelled as algorithmic) [42].

H. Functionality and Security

There is a tacit assumption and, in some quarters, misplaced trust that the functionality and security of devices monotonically improves over time with each new model, iteration or software update. Some updates and version changes, particularly those making significant changes to operating systems, popular apps or device behaviors, can provoke strong user responses as, for example, evidenced by Microsoft's Windows 8 [43] and Apple's iPhone 'Batterygate' [44]. But, more often, updates make bug fixes and iterative adjustments to algorithms, interfaces and features [45] and have little immediate, if any, observable effects, and there is a tacit assumption amongst users that these iterations generally evolve systems over time into better-performing, more functional, secure and efficient versions of their previous selves. This may often be the case, but it is not necessarily true, as for example demonstrated by the 'Dieselgate' automotive emissions scandal [46].

IV. DISCUSSION AND CONCLUSIONS

A. Fairness Drivers and Barriers Mismatch

A key observation when contrasting fair technology drivers with fair technology barriers is that the drivers do not compound to increase fairness in the same way that the barriers compound to limit fairness. Drivers generally necessitate compliance with *minimum* standards and there is, on balance, a lack of incentives for technology enterprises (and stakeholders) to further enhance fairness. This mismatch between compounding monotonically decreasing fairness barriers and 'minimum requirement drivers' means that unfairness will be programmed into the digital ecosystem unless drivers with the capacity to monotonically increase fairness can be effected.

B. Future Challenges

There are clearly open challenges and a need for further research and empirical evidence but, in terms of progress, there is evidence of positive development in terms of the evolution of standards, tools and assurance solutions being developed to mitigate against algorithmic bias. At the same time there is a need to implement meaningful 'human-in-the-loop' protocols to ensure algorithmic and digital technologies augment human decision-making, minimize unfair outcomes and build trust in the wider operation of such systems. Such solutions are complex and require utilizing skills from the Human-Computer Interaction (HCI) and User Experience (UX) fields plus wider multi-disciplinary inputs including cognitive science, anthropology, moral philosophy, social sciences and law, to establish a human-centered AI approach [47]. In this way we envision a new inter-disciplinary field of 'Human-AI interaction', that encompasses Human-Computer Interaction and the engineering equivalent of Human-Machine Interaction (HMI) and specifically brings, with the intention not only to inform design frameworks for widespread adoption and interaction with smart technology, but also legal frameworks as well.

REFERENCES

[1] J. Rawls, "Justice as Fairness: Political not Metaphysical," in Philosophy and Public Affairs, 14(3), pp. 223–251, 1985.

[2] R. K. Jain, D. M. W. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination," Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA, 1984.

[3] H. Vandierendonck and A. Seznec, "Fairness Metrics for Multi-Threaded Processors," in IEEE Computer Architecture Letters, vol. 10, no. 1, pp. 4-7, 2011.

[4] S. H. I. Huaizhou, R. V. Prasad, E. Onur, and I. G. M. M. Niemegeers, "Fairness in wireless networks: Issues, measures and challenges," in IEEE Communications Surveys & Tutorials, 16(1), pp. 5-24, 2013.

[5] J. Pitt, "Interactional justice and self-governance of open self-organising systems," in 2017 IEEE 11th International Conference on Self-Adaptive and Self-Organizing Systems (SASO), pp. 31-40, 2017

[6] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. Daumé III, and K. Crawford, "Datasheets for datasets," arXiv preprint arXiv:1803.09010, 2018.

[7] E. Ntoutsi, P. Fafalios, U. Gadiraju, V. Iosifidis, W. Nejdl, M.E. Vidal, S. Ruggieri, F. Turini, S. Papadopoulos, E. Krasanakis, and I. Kompatsiaris, "Bias in data-driven artificial intelligence systems - an introductory survey," in Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), p.e1356, 2020.

[8] K. Sandler, L. Ohrstrom, L. Moy, and R. McVay, "Killed by code: Software transparency in implantable medical devices," Software Freedom Law Center, pp. 308-319, 2010.

[9] L. M. Cysneiros, M. Raffi, and J. C. Sampaio do Prado Leite, "Software transparency as a key requirement for self-driving cars," in IEEE 26th International Requirements Engineering Conference (RE), Banff, Alberta, Canada, pp. 382-387, 20th - 24th August, 2018.

[10] J. Herkert, J. Borenstein, and K. Miller, "The Boeing 737 MAX: lessons for engineering ethics," in Science and Engineering Ethics, 26(6), pp. 2957-2974, 2020.

[11] G. Travis, "How the Boeing 737 Max disaster looks to a software developer," in IEEE Spectrum, 18, https://spectrum.ieee.org/aerospace/aviation/how-the-boeing-737-max-disaster-looks-to-a-software-developer (retrieved May 2021), 2019.

[12] J. Christie, "The Post Office Horizon IT scandal and the presumption of the dependability of computer evidence," in Digital Evidence & Elec. Signature L. Rev., 17, p. 49, 2020.

[13] H. Dixon, P. Sawer, and T. Diver, "Call to prosecute Post Office bosses over 'biggest miscarriage in British legal history'," The Telegraph, 23rd April, 2021.

[14] K. Peachey, "Post Office scandal: What the Horizon saga is all about," BBC News, 23rd April, 2021.

[15] J. M. Haney, and S. M. Furman, "Smart Home Updates: User Perceptions and Experiences," in Symposium on Usable Privacy and Security (SOUPS), Boston, MA, USA, 9th - 11th August, 2020.

[16] J. Pitt and J. Ober, "Democracy by Design: Basic Democracy and the Self-Organisation of Collective Governance," in 2018 IEEE 12th International Conference on Self-Adaptive and Self-Organizing Systems (SASO), pp. 20-29, 2018.

[17] J. Pitt, "From trust and loyalty to lock-in and digital dependence," in IEEE Technology and Society Magazine, 39(1):5-8, 2020.

[18] C. Criado-Perez, Invisible women: exposing data bias in a world designed by men, Vintage, London.(Trust), 2020.

[19] A. J. McGregor, Sex Matters: How male-centric medicine endangers women's health and what we can do about it, Quercus, London, 2020.

[20] P. Y. Chau, "An empirical assessment of a modified technology acceptance model," in Journal of management information systems, 13(2), pp. 185-204, 1996.

[21] T. Collins, S. I. Woolley, S. Oniani, I. M. Pires, N. M. Garcia, S. J. Ledger, and A. Pandyan, "Version reporting and assessment approaches for new and updated activity and heart rate monitors," in Sensors, 19(7):1705, 2019.

[22] T. Collins, and S. I. Woolley, "Some heart-rate monitors give less reliable readings for people of colour," in The Conversation, Health + Medicine, 1st August, 2019.

[23] S. Oniani, S. I. Woolley, I. M. Pires, N. M. Garcia, T. Collins, S. Ledger, and A. Pandyan, "Reliability assessment of new and updated consumer-grade activity and heart rate monitors," in Sensor Devices, Venice, Italy, 16th - 20th September, 2018.

[24] B. Bent, B. A. Goldstein, W. A. Kibbe, and J. P. Dunn, "Investigating sources of inaccuracy in wearable optical heart rate sensors," in NPJ Digital Medicine, 3(18), 1-9, 2020.

[25] P. J. Colvonen, P. N. DeYoung, N. O. A. Bosompra, and R. L. Owens, "Limiting racial disparities and bias for wearable devices in health science research," in Sleep, 43(10), zsaa159, 2020.

[26] T. Rukasha, S. I. Woolley, T. Kyriacou, and T. Collins, "Evaluation of wearable electronics for epilepsy: A systematic review," in Electronics, June 9(6):968, 2020.

[27] S. Krüger, and N. Ni Bhroin, "Vital signs: Innovations in self-tracking health insurance and social change," in Journal of Media Innovations, 6(1), 2020.

[28] L. Giddens, D. Leidner, and E. Gonzalez, "The role of Fitbits in corporate wellness programs: Does step count matter?" in 50th Hawaii International Conference on System Sciences, pp. 3658-3667, Hawaii, USA, 4th - 7th January, 2017.

[29] D. Hankerson, A. R. Marshall, J. Booker, H. El Mimouni, I. Walker, and J. A. Rode, "Does technology have race?" in CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 473-486, San Jose, CA, USA, 7th – 12th May, 2016.

[30] J. Galbally, R. Haraksim, and L. Beslay, "A study of age and ageing in fingerprint biometrics," in IEEE Transactions on Information Forensics and Security, 14(5), pp. 1351-1365, 2018.

[31] T. Rukasha, S. I. Woolley, and T. Collins, "Wearable epilepsy seizure monitor user interface evaluation: an evaluation of the Empatica 'Embrace' interface," in Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, pp. 110-114, 2020.

[32] N. W. Moon, P. M. Baker, and K. Goughnour, "Designing wearable technologies for users with disabilities: Accessibility, usability, and connectivity factors," in Journal of Rehabilitation and Assistive Technologies Engineering, vol. 6, 2019.

[33] C. O'Neil, Weapons of math destruction: How big data increases inequality and threatens democracy, Crown, New York, 2016.

[34] A. L. Smith, A. Chaudhuri, A. Gardner, L. Gu, M. B. Salem, M. Lévesque, "Regulatory frameworks relating to data privacy and algorithmic decision making in the context of emerging standards on algorithmic bias," NIPS Conference Workshop on Ethical, Social and Governance Issues in AI, Montreal, Canada, 7th December, 2018.

[35] D. Danks, and A. J. London, "Algorithmic Bias in Autonomous Systems," IJCAI, vol. 17, pp. 4691-4697, 2017.

[36] P. Schmidt, F. Beissman, and T. Teubner, "Transparency and trust in artificial intelligent systems," in Journal of Decision Systems, 29(4), pp. 260-278, 2019.

[37] M. Kearns, and A. Roth, The ethical algorithm: The science of socially aware algorithm design, Oxford University Press, 2019.

[38] J. Ballance, R. Clifford, and D. Shoag, "No more credit score: Employer credit check bans and signal substitution," in Labour Economics 63, 2020.

[39] J. Buolamwini, and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in Conference on fairness, accountability and transparency, pp. 77-91, PMLR, 2018.

[40] K. Kirkpatrick, "Battling algorithmic bias: how do we ensure algorithms treat us fairly?" in Communications of the ACM 59, no. 10: 16-17, 2016.

[41] A. Gardner, A. L. Smith, A. Steventon, E. Coughlan, and M. Oldfield, "Ethical funding for trustworthy AI: proposals to address the responsibilities of funders to ensure that projects adhere to trustworthy AI practice," in AI Ethics, pp. 1-15, 2021.

[42] E. Bogert, A. Schecter, and R. T. Watson, "Humans rely more on algorithms than social influence as a task becomes more difficult," in Scientific Reports, 11, 8028, 2021.

[43] N. Jansen, "Usability evaluation of Windows 8 with keyboard and mouse: challenges related to operating system migration in large organizations," Master's thesis, Institutt for datateknikk og informasjonsvitenskap, 2013.

[44] H. J. Lee, K. H. Lee, and J. Choi, "A structural model for unity of experience: connecting user experience, customer experience, and brand experience," in Journal of Usability Studies, 11(1), 2018.

[45] S. I. Woolley, T. Collins, J. Mitchell, and D. Fredericks, "Investigation of wearable health tracker version updates," in BMJ health & care informatics, 26(1):e100083, 2019.

[46] L. Bovens, "The ethics of Dieselgate," in Midwest Studies in Philosophy, 40, pp. 262-283, 2016

[47] R. R. Bond, M. Mulvenna, D. Finlay, A. Wong, A. Koene, R. Brisk, J. Boger, and T. Adel, "Human centred artificial intelligence: weaving UX into algorithmic decision making," in Proceedings of RoCHI, Bucharest, Romania, 17th-18th, pp. 2-9, October, 2019.