University of Birmingham

Towards an equitable digital society

Elliott, Karen; Price, Rob; Shaw, Patricia; Spiliotopoulos, Tasos; Ng, Magdalene; Coopamootoo, Kovila; van Moorsel, Aad

[Link to publication on Research at Birmingham portal](#)

ORIGINAL ARTICLE

Towards an Equitable Digital Society: Artificial Intelligence (AI) and Corporate Digital Responsibility (CDR)

Karen Elliott[1] · Rob Price[2] · Patricia Shaw[3] · Tasos Spiliotopoulos[1] · Magdalene Ng[1] · Kovila Coopamootoo[1] · Aad van Moorsel[1]

Abstract

In the digital era, we witness the increasing use of artificial intelligence (AI) to solve problems, while improving productivity and efficiency. Yet, inevitably costs are involved with delegating power to algorithmically based systems, some of whose workings are opaque and unobservable and thus termed the "black box". Central to understanding the "black box" is to acknowledge that the algorithm is not mendaciously undertaking this action; it is simply using the recombination afforded to scaled computable machine learning algorithms. But an algorithm with arbitrary precision can easily reconstruct those characteristics and make life-changing decisions, particularly in financial services (credit scoring, risk assessment, etc.), and it could be difficult to reconstruct, if this was done in a fair manner reflecting the values of society. If we permit AI to make life-changing decisions, what are the opportunity costs, data trade-offs, and implications for social, economic, technical, legal, and environmental systems? We find that over 160 ethical AI principles exist, advocating organisations to act responsibly to avoid causing digital societal harms. This maelstrom of guidance, none of which is compulsory, serves to confuse, as opposed to guide. We need to think carefully about how we implement these algorithms, the delegation of decisions and data usage, in the absence of human oversight and AI governance. The paper seeks to harmonise and align approaches, illustrating the opportunities and threats of AI, while raising awareness of Corporate Digital Responsibility (CDR) as a potential collaborative mechanism to demystify governance complexity and to establish an equitable digital society.

Keywords Artificial intelligence (AI) governance · Digital ethics and trust · Complexity · Corporate Digital Responsibility · Equitable digital society · Financial technology (FinTech)

Introduction

The global financial crisis (GFC, 2007–2009) marked a significant failure in citizens' trust and questioned governance mechanisms in financial services (Pedersen 2021). Several algorithms ceased working during the crisis because of the degree of stress and strain on the calibration designed in a different epoch, a calibration that lacked the ability to adapt the behaviour of the algorithm under a crisis situation, creating a breakdown of societal trust in artificial intelligence (AI) (Edelman 2019a, 2019b). Such concerns are revealed in public perceptions and uncertainty surrounding AI's future in

✉ Karen Elliott
karen.elliott@newcastle.ac.uk

Rob Price
info@corporatedigitalresponsibility.co.uk

Patricia Shaw
trish@beyondreach.uk.com

Tasos Spiliotopoulos
tasos.spiliotopoulos@newcastle.ac.uk

Magdalene Ng
magdalene.ng@newcastle.ac.uk

Kovila Coopamootoo
kovila.coopamootoo@newcastle.ac.uk

Aad van Moorsel
aad.vanmoorsel@newcastle.ac.uk

[1] School of Computing & Business School, Newcastle University, 1 Science Square, The Helix, Newcastle upon Tyne NE4 5TG, UK

[2] http://CorporateDigitalResponsibility.net (CDR), Alchemmy, 52-54 High Holborn, London WC1V 6RL, UK

[3] Beyond Reach Consulting Ltd, 139 Furlong Road, Bolton-Upon-Dearne, Rotherham S63 8HD, UK

society from technology executives overseeing the development and implementation of AI to the general public (in the USA). Key findings underpin the central tenets of the field of "AI and Society", which this paper examines. On the one hand, industry and people are curious about "good" AI—the opportunities and benefits. On the other hand, we observe substantial uncertainty and anxiety that the rapid adoption of AI across the digital "space" will impact society in negative ways: widespread job loss, income inequality, and social isolation (ibid.). How can citizens be included and benefit from AI innovations and, specifically, protect their data and digital identities in the fourth industrial revolution (Kowalikova et al. 2020; Lewis 2021; Hamdan et al. 2021)? And, where does accountable responsibility rest to ensure that we establish an equitable digital society for all, and not the few? To place the debate on the associated risks of AI and societal implications in context, first, we review pertinent statistical insights in this area; second, we provide a brief background to our current financial services (FS) and financial technology (FinTech) environment, specifically, AI/FinTech-enabled FS; third, we examine trust, ethical principles, and regulatory implications (including the new EU guidance). Last, we introduce Corporate Digital Responsibility (CDR) as a potential collaborative mechanism to navigate such complexity, proposing guidance frameworks towards responsible corporate digital actions in preserving societal interests.

AI Perceptions

Edelman's AI survey (2019b) findings raise fundamental questions for AI/FinTech-enabled FS adoption and associated societal implications. The public reported slightly higher AI concerns than the tech executives canvassed in the USA (percentages indicated respectively); *safe development*—60/54%; *the roles of society, business, and government*—91/84%; *hurting the poor*—54/43%; *benefiting the wealthy*—67/75%; *loss of human intellectual capabilities*—71/65%; *increased social isolation*—74/72%; *society feels threatened*—81/77%; and *highly corrosive to public trust*—51/45%. Each finding highlights the interplay between perceived AI risks, opportunities, and threats that are interwoven within our discussion surrounding notions of trust, ethics, legality, and governance nested in societal systems and people's relationship with AI. Simply put, technologies have permeated social interaction whether on digital platforms or the positives afforded by Zoom during the recent pandemic (Wiederhold 2020; Haochen and Polak 2021). Online interactions are imbued with decisions around our levels of awareness regarding privacy and security of our data, in readily accepting the "T&Cs" (terms and conditions), or General Data Protection Regulation (GDPR 2018) compliance via the click of a digital button to access information we desire. The technology that lies

beneath, regulating and accessing a range of complex systems, that can arbitrarily create a characteristic profile of our digital "selves" and AI continues to develop.

AI Development

Lepore (2020: 2) names the Simulmatics Corporation as first to draw on Turing's insights and "engage principally in estimating probable human behaviour by the use of computer technology." Simulmatics is "[t]he long-dead grandparents of the data-mad, algorithmic twenty-first century" (ibid: 4-5). This corporation employed the "What-If Men" who forged ahead on the assumption that in replicating human behaviour, many societal disasters and risks could be averted via the use of technology. As a result, this group instigated the "future computed" as AI now dominates society's symbiotic existence (Microsoft 2018). Moving forward, Cybenko's (1989) "Universal approximation theorem" proof showed that artificial neural networks (ANNs) can, with arbitrary precision, approximate all continuous functions with a finite number of learning nodes. This is a powerful feature of AI-based algorithms. For instance, distributed ledger technology (DLT) that emerged in 2008 employs this powerful feature, providing multiparty computations wide-scale transaction clearing with mutual distrustful parties, termed FinTech, opening a universe of AI/FinTech-enabled financial services (FS). A recent use-case leveraging DLTs/AI/FinTech-enabled FS is open banking (OB). Customers are given access to their own data, previously the reserve of traditional financial institutions and data services "records of everything upon which a customer makes an electronic payment", permitting products and services tailored to the fluctuating financial habits and demands of the tech "savvy" customer (Bell 2020: 183). As we shall reveal, a feature of AI-governance is that rather than "compliance" with AI regulation or statutes of law deemed necessary, "doing good business" becomes a cultural "norm" exemplified within OB practices to benefit society (Durodié 2019: 121).

This overview demonstrates AI's potential for "good" societal interventions—from increased efficiency in analysing large datasets, reduction of mundane tasks with reliability and consistency, opening financial services, to complementing hazardous human tasks—bomb disposal, and precise medical interventions (Maddeo and Floridi 2018). Given the "good" tempered with the caveat of "bad" public opinion in the Edelman (2019a, 2019b) survey, and the reported mistrust of AI systems to behave in accordance with its intended purpose, do we know that AI engineering decisions are unbiased and fair and promote equality of use aligned with societal values? (Aitken et al. 2020). A significant issue surrounds the "black-box" element of AI whereby computer scientists and engineers find that the machine learning systems used for predicting an outcome remain opaque or

cannot be adequately observed. Granted, a level of acceptance can be discerned to satisfy machine learning principles, but progress in trying to resolve these issues will be a feature of auditability and understanding of AI decisions ex post to enhance broader uptake of machine learning and thus, AI (Pasquale 2015, 2017). Explaining the internal mechanics of the deep or machine learning system in human terms or explainable AI (XAI) remains problematic because there is "[n]o real consensus about what interpretability is in machine learning" (Molnar 2019: 31). Hence, the XAI body of literature is nascent and may take several years to explore, understand, and translate into a useable format (Elton 2020). Meanwhile, regulators and the law attempt to address the risks and culpability of the "black box" phenomenon of AI systems, decisions, and subsequent societal effects (Pasquale 2015, 2017). What is required is to move "from more AI to better AI" (Pasquale 2020: 196).

At the time of writing, nothing is compulsory in terms of regulation, AI governance, ethics, and legal compliance trail the rate of AI innovation and implementation in digital society (Floridi 2019; Roitblat 2020). Indeed, as Mittelstadt (2019: 501) asserts, "ethical principles are not enough." Certain flaws exist in the "84 public-private initiatives" he explored regarding adequately defined duties, accountability, linkage to societal norms, and a framework of methods for AI implementation and governance. Over 160 ethical principles exist to date, suggesting fragmentation and difficulty in choosing which principles are best for responsible use of AI (AlgorithmWatch 2020). The European Union consulted across academia, industry, and policymakers to agree upon a set of co-created standards for AI Regulation—released on 21 April 2021 (EU, 2021)[1]. Yet, early insights indicate that interpretation of the EU rules, compliance, and enforcement are "vague", and "loopholes" are already identified in the draft documentation (Bloomberg 2021; The Verge 2021); we will return to this regulation later. Now, we move to examine the digital society landscape and its social actors.

Digital Society

Zuboff (2019) examined digital society, AI, and symbiotic social relationships, claiming that society is subject to *Surveillance Capitalism*. Technologies track our every movement, and almost unaware, we have slipped into accepting digital surveillance as a daily norm. Despite adopting a predominant Western perspective, Zuboff's (2019: 199-232) work draws attention to the discourse we unpack; she reveals the notion of overt and covert data manipulation, whereby users have a "puppet and puppet master" relationship (ibid.:

14-17)—the puppet representing a device to grant access to the digital society, which has AI systems operating in the background, harvesting data insights and forming arbitrary user profiles aligned to the masters' interests (cf. Cybenko 1989). She argues this process is enabled via the "forward text" and the "shadow text". The former is the user experience (UX), to be aware of our data and who is using said data, when engaged online via social media platforms, i.e. Google and Facebook. UX is premised on an array of alluring features to "hook" and retain user's attention and stimulate their desire for continued engagement. Conversely, the "shadow" text, as the term suggests, describes a covert system owned and manipulated to benefit the master, via sharing and monetising user's data with selected third parties. Transparency to the user and society is blurred, for the tech giants operate largely without regulation, using and re-using data (ibid.). A moot point is the exploitation of adolescence, "[y]oung life now unfolds in the spaces of private capital, owned, and operated by surveillance capitalists…operationalised in practices designed to maximise surveillance revenues" (Zuboff 2019: 456). Adulthood mirrors this description, as the puppet master lures individuals to join the digital society (ibid.: 335):

> the machine intelligence capabilities to which data are continuously supplied, the analytics that discern patterns, and the algorithms that overt them into rules . . . the essence of the uncontract, which transforms the human, legal and economic risks of contracts into plans constructed, monitored and maintained by private firms for the sake of guaranteed outcomes: less contract utopia than *uncontract dystopia*

Translating the uncontract for AI/FinTech-enabled FS, we suggest that beyond OB, society has been changed, premised on the new "normative judgment", that the digital society is a better state than the prior (Byrne 2010: 62). Byrne (ibid.) draws our attention to the pivotal ethical question involving the normative self and how the plethora of actors within AI/FinTech-enabled FS interact: we "assume that "we" represent some sort of universal interest—in reality there are often conflicting interests at play—what works for whom?"

Vested Interests

Zuboff's findings exemplify conflicting interests between the *puppet master* orchestrating the *uncontract dystopia*, reifying Alford's (1975: xiii) notion of "strategically structured interest" first applied to examine the conflicting interests in health system reforms. Alford contended there existed "a continuing struggle between major structural interests operating within the context of a market society – "professional monopolists"

controlling the major . . . resources, "corporate rationalisers" challenging their power, and the community population seeking better . . . care" (ibid. : xiv). Based on the above, we interpret the current digital society in Fig. 1.

Figure 1 echoes Zuboff's (2019) insights. The monopolist position is the *shadow text* level. The invitation of the monopolisers to the community population to participate in the digital society amounts to "social control"—the *puppet* (Alford, and R.,, and Friedland, R. 1975: 472) or Zuboff's (2019: 334-6) "blankness of perpetual compliance". Hence, analogies between Zuboff's appeal to maintain human qualities of dialogue, problem-solving, and empathy against the "uncontract" environment are juxtaposed against a perceived ineffectual meeting of non-dominant needs and subsequent "low-levels of engagement", particularly amongst "lower-income individuals" linking to digital inequity of the "community population" residing in the digital divide (ibid.; van Dijk, 2020). We return to discuss the corporate rationalisers or *forward text* role, later in this paper. Alford's theory has been criticised for not defining which interests are specifically "repressed" by the monopolisers (North and Peckham 2001; Checkland et al. 2009). Nonetheless, the approach is useful and continues to be applied across finance, and to COVID-19, relative to conflicting interests (Ülgen 2017; Briggs 2021). We have acknowledged conflicting interests residing in the digital society, binding us to the "normative or ethical domain . . . Normative issues are intertwined with our very understanding of complexity" (Cilliers 2005: 259-264). In order to make responsible decisions for AI/FinTech-enabled FS and society, we must respect "otherness and difference as values in themselves" (Cilliers 1998: 139). To echo Kunneman (2010: 132), addressing "complexity is not only the central scientific, but also the central *ethical* problem of our time". We know that algorithms can reconstruct people's data and discriminate around gender and semantics (Perez 2019; Toreini et al. 2020). However, the subjective concept of "trustworthiness" to remove bias is difficult in computer science

engineering discourse and practice; hence, we consider the complexities in defining and building trust, in light of the plethora of ethical principles in circulation.

Trust, Ethics and Human Oversight of AI

Koshiyama et al. (2021: 2) recognise a series of sociotechnical and ethical issues characterised by the shift from "Big Data" to "Big Algo" in the "5 V" model: (i) *volume*, as resources and know-how proliferate, soon there will be billions of algorithms; (ii) *velocity*, algorithms making real-time decisions with minimal human intervention; (iii) *variety*, from autonomous vehicles to medical treatment, employment, finance, etc.; (iv) *veracity*, reliability, legality, fairness, accuracy, and regulatory compliance as critical features; and (v) *value*, new services, sources of revenue, cost-savings, and industries will be established. Thus, exposing the complexities of definition, debates, opaqueness and what the future holds, nobody has a complete grasp on the full potential of machine learning and AI not to mention its oversight (Hauer 2018, 2019). On the one hand, COVID-19 highlights the increased rate of digital skills, adoption, and transformation across digital society, viewed as a positive outcome of the pandemic ("volume"; Iivari et al. 2020). On the other hand, this trend fails to include the marginalised in society languishing in the digital divide and beholden to the *puppet-puppet master* scenario (van Dijk 2019). Thus, we wrestle with the consequences: when the algorithm is "responsible" for a societal harm, does this align with our societal "veracity" (Abbott 2018; Shaw 2021). Bryson (2021) likens the widespread adoption of AI to "electricity . . .[or] nuclear technology" where variants of the "good" versus "bad" debate weigh heavily on our normative ethical and moral instincts—to ignore the bad in favour of the good aspects of AI would be foolhardy given Zuboff's (2019) insights. Likewise, we cannot ignore vested interests in advocating and luring society into technological adoption

Fig. 1 Structured vested interests in the digital society (cf. Alford 1975)
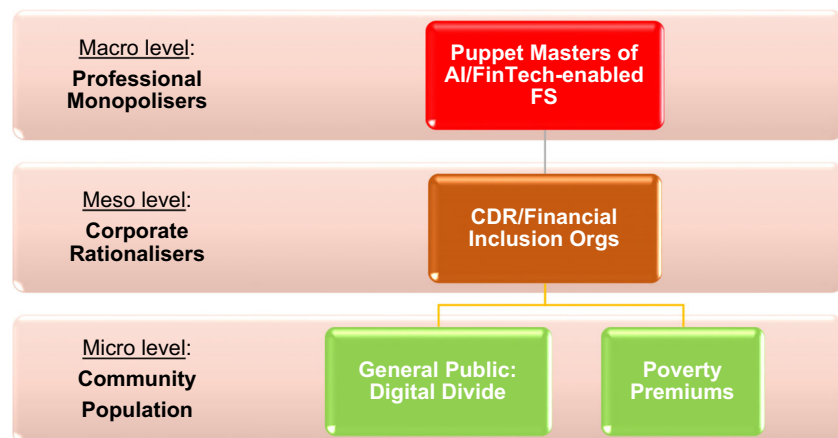
FINANCIAL SERVICES REMAINS LEAST TRUSTED

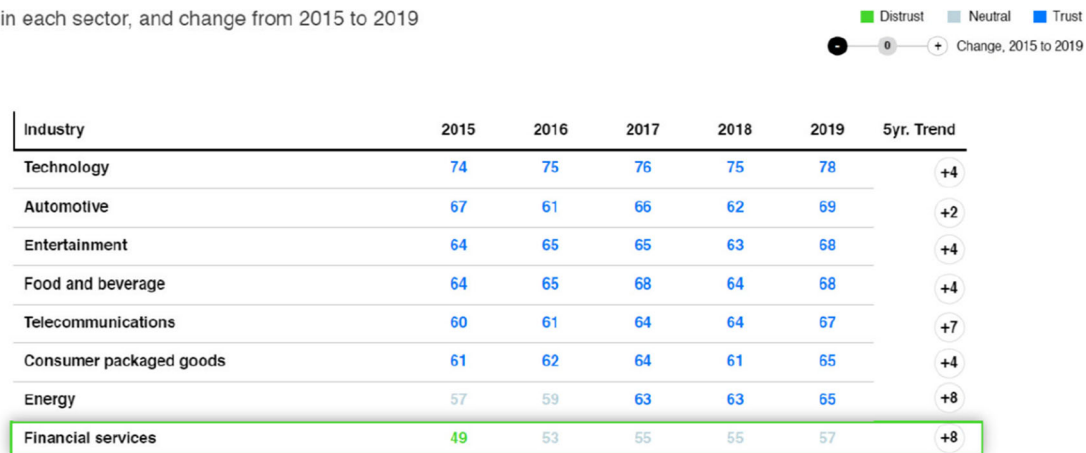Trust in each sector, and change from 2015 to 2019

Distrust   Neutral   Trust

●———— 0 ————+ Change, 2015 to 2019

Industry	2015	2016	2017	2018	2019	5yr. Trend
Technology	74	75	76	75	78	+4
Automotive	67	61	66	62	69	+2
Entertainment	64	65	65	63	68	+4
Food and beverage	64	65	68	64	68	+4
Telecommunications	60	61	64	64	67	+7
Consumer packaged goods	61	62	64	61	65	+4
Energy	57	59	63	63	65	+8
Financial services	49	53	55	55	57	+8

Fig. 2 Trust in financial services 2015–2019 (source: Edelman 2019b: 8)

("variety" and "value"). Mitchell (2019: 145) warns, "the science of electricity was well understood before it was widely commercialized. We are good at predicting the behaviour of electricity . . . [such is] not the case for many of today's AI-systems" "velocity". If we want AI/FinTech-enabled FS to be trusted, combined with overcoming the above issues and ramifications post-GFC (2007-9), work is required to re-build trust, given this sector's current "least trusted" label awarded by its customers (Edelman 2019b; see Fig. 2).

AI's impact for an equitable digital society is not inevitable (Pasquale 2020). Progress towards equity must be intentional, coordinated, and, crucially, *collaborative*. Trustworthy outcomes in AI/FinTech-enabled FS require institutions and organisations to take a holistic approach to the management and oversight of not just financial products, but AI (and the data that empowers it) to engender trust. Trust is not reliant on the AI, or the outcomes it produces; rather, trust in an organisation's ability to create responsible AI. Despite many traditional financial institutions possessing a social licence to operate, consumer trust is fluid, must be earned, maintained, and as we saw with the GFC (2007-9; Table 1) can easily be lost (Dietz and Gillespie 2012; Aitken et al. 2020).

TRUST requires transparency, responsibility, understanding, stewardship, and truth (Shaw 2020: 176, italics and brackets added):

Transparency being clear about what it is you are doing and why . . . the opportunities and risks. . . *Responsibility* being reputable and accountable for what it is you do; *Understanding* . . . you provide [customer] services to understand what outcome they can expect and how it will impact them...and society; *Stewardship*...a good custodian of the data . . in line with the kind of society we all want to create; *Truth,*

validating the accuracy of data, the insight and inferences made . . . are beneficial and not harmful.

TRUST must be codified into practical ethical data and AI governance tools, which are not mutually exclusive but understood within an overlapping framework (i.e. AI governance), if we are to innovate for socially beneficial AI/FinTech-enabled FS, worthy of our shared digital futures. Earlier, we referred to the Draft EU AI Regulation (DEAR) published on 21st April 2021.[2] It delineates a risk-based approach to AI, imposing organisational-level governance in respect of data and quality management procedures and ongoing monitoring, including transparency and provision of information for users[3] across the AI lifecycle, only vis-a-vis "High-risk AI."[4] DEAR is only applicable to the EU-centric AI marketplace, rendering FinTechs outside the EU at a disadvantage, suggesting a lower bar for trust, vis-a-vis EU comparators who want to operate globally. Although DEAR progresses towards providing regulatory impetus for AI governance, giving organisations legitimacy to ascribe budget and costs (i.e. ethical implementation requirements), the guidance permits high-risk AI operators to decide the method/technical solution to achieve compliance. DEAR implementation will also be challenging for EU market-entrant FinTech start-ups compared to larger established competitors, as DEAR leans heavily on governance, market monitoring, and enforcement roles. We argue that embedding TRUST becomes a pivotal method of competitive advantage (Arnold et al. 2019).

---

[2] https://eur-lex.europa.eu/procedure/FI/2021_106

[3] "Users" defined in DRAFT Article 3(4) as "'user' means any natural or legal person, public authority, agency or other body using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity", not end users.
[4] Draft Art 6 and Annex III, DRAFT EU AI Regulation supra.

Similarly, DEAR affirms the significant void for "cuspy",[5] medium- and low-risk AI, devoid of regulatory guidance and practical tools to implement effective AI governance, reliant on voluntary industry, or organisational codes of conduct[6] to embed them. Thus, AI governance remains in a vacuum, open to the five unethical risks: (1) ethics shopping; (2) ethics blue-washing; (3) ethics lobbying; (4) ethics dumping; and (5) ethics shirking, undermining the potential to trust the use of AI (see Floridi 2019). To engender trust, DEAR calls for standards, conformity assessment, and certification, an end-user facing Conformité Européene (CE) Marking, and registration with a public EU Database for High-Risk AI. However, precise administrative details for compliance remain unclear. DEAR could have collaborated with existing ethical AI standards with shared objectives in progressing to maturity of in-organisation operational processes. For example, the Institute of Electrical and Electronics Engineers (IEEE)[7] and International Organization for Standardization (ISO)[8] provide "AI standardization . . . to establish trust in AI-systems", specifically, an AI playbook[9] for financial services "to prioritise human well-being and ethical considerations." While DEAR acknowledges "[a] comprehensive ex-ante conformity assessment through internal checks, combined with a strong ex-post enforcement, could be an effective [solution]"[10] to promote public trust, DEAR also recognises the nascent status of such processes, including that "expertise for [AI] auditing is only now being accumulated" (ibid.). For instance, ForHumanity (2021)[11] lead this area in creating an "infrastructure of trust", through AI-auditing standards and tools that have the potential to impact humans in the areas of "Ethics, Bias, Privacy, Trust and Cybersecurity". We can infer that these ex ante and ex post governance tools are intended to operate in providing transparency while engaging a "soft-law" normative constraint on "the "do's" and "don'ts" of algorithmic use in society" (Morley et al. 2020); without enforceability or liability, these "soft law" tools lack "teeth" (Shaw 2021),

Nonetheless, DEAR marks a significant improvement for AI governance, affording organisations the opportunity to find and create risk management, governance, and oversight solutions provided conformity is achieved. Conversely, DEAR risks invoking a "tick-box" compliance culture (FCA 2016; van Vuuren 2020), rather than espousing a digital ethics culture across the AI lifecycle and digital society (see CDR below). This does not compensate for the lack of real and

meaningful sociotechnical interaction between internal and external "end user" stakeholders. DEAR underplays the importance of co-governance in stakeholder engagement to de-risk AI and hold it accountable to build trust (Ackerman 2004), in respect of bias, ethical and societal impacts, which ultimately lead to legal consequences for AI systems businesses (Coeckelbergh 2020; Toreini et al. 2020). Such core elements are the heart of outcome-based risks associated with AI systems, which can and do undermine trust. Pivotally, "external ethical auditing will be a key component of any form of operationalised AI-ethics" (Morley et al. 2021: 11), where "individuals have a right to expect the technology they use to perform in a reasonable manner and to respect their trust"[12]. Without trust, social responsibility and operationalising digital ethics through AI-governance, regulation will fail as "consumers won't use the firm's services, adopt new technology, or share their data" (Shaw 2020: 176). DEAR provides insufficient methods of "how" to be digitally responsible for AI/FinTech-enabled FS; hence, collective groups are co-creating mechanisms to address responsibility.[13]

Corporate Digital Responsibility (CDR)

CDR is a voluntary commitment by organisations fulfilling the *corporate rationalisers'* role in representing community interests to inform "good" digital corporate actions and digital sustainability (i.e. data and algorithms) via collaborative guidance on addressing social, economic, and ecological impacts on digital society. For AI/FinTech-enabled FS, the CDR is a potential framework to assist navigating AI governance complexity and to devise an informed strategy. In short, AI governance post-GFC must ensure equity beyond the monopoliser's interests to include all stakeholders invested in an organisation's modus operandi (Bell 2020; Pedersen 2021). CDR codifies TRUST and illustrates how AI governance and expectations are met building on lessons learned from corporate social responsibility including environmental and sustainable goals (CSR, Carroll 1991). Limited literature exists, but we advocate CDR as a separate proposition for organisations specifically linked to digital technology and data (Lobschat et al. 2021), not an extension of CSR (Herden et al. 2021). Rather, CDR complements such protocols, as the digital realisation of the same responsibilities but with a two-fold appropriate use of digital and data *within* and *by* the organisation to impact society while sustaining our planet to improve organisational environmental and social governance performance (Dörr 2020). Thus, the combination of environmental and social responsibilities is viewed as the transparent measurement of effectiveness in CDR execution, while accessible for

---

[5] AI-systems, on the cusp of the high-risk AI bracket, open to judicial/administrative interpretation.
[6] DRAFT Article 69.
[7] https://ethicsinaction.ieee.org/p7000/
[8] https://www.iso.org/committee/6794475.html
[9] https://ethicsinaction.ieee.org/#series
[10] Draft Art 6 and Annex III, DRAFT EU AI Regulation supra.
[11] US-based, non-profit organisation, https://forhumanity.center/independent-audit-of-ai-systems.

[12] https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html
[13] https://www.itechlaw.org/ResponsibleAI2021

Fig. 3 Corporate Digital Responsibility/Digital Responsibility Code (DRC) copyright© CDR.net (https://corporatedigitalresponsibility.co.uk/—draft version, CDR group working on developing this graphic.)
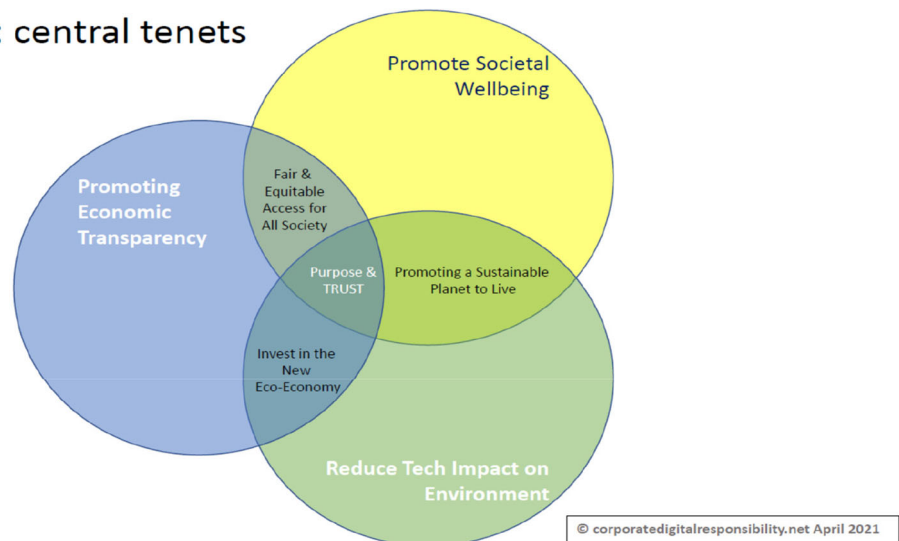
stakeholders to evaluate organisational performance operating within digital society (Bell 2020: 201), illustrated in Fig. 3.

Interpreting Fig. 3 for AI/FinTech-enabled FS governance, "purpose and trust" are the central drivers for the appropriate use of digital and data *within* and *by* the organisation(s). Trust is subjective; therefore, a "declaration of purpose" by organisations is advocated, committing to upholding the collective interests of society and the planet via innovative, appropriate, and permissible use of technologies (Dietz and Den Hartog 2006; Dietz and Gillespie 2012). For example, adhering to a *Digital Responsibility Code* (*DRC*), linked to the notion of green finance, aligned with Sustainable Development Goals (SDGs) across AI lifecycles (Lindenberg 2014, Taghizadeh-Hesarya and Yoshinob 2019). Data efficiency is driven in reducing consumption and measuring emissions using digital technologies to offset carbon impact while investing in environmental, ecotech, and cleantech digital solutions. Likewise, there is a need to embrace tenets of "Fairer access for All" to allay societal fears of AI automation and job loss for governance to succeed. Hence, "good" uses of AI and the economic benefits of digitisation (such as taxation) should be evident to the public (Pasquale 2020). Similarly, responsible engagement with the gig economy (where appropriate) is a must, and abiding by legal, regulatory, and ethical principles per geography and market, and embodying an "open" data approach to demonstrate commitment to the DRC (Shaw 2020, 2021). In so doing, an AI/FinTech-enabled FS can embrace the tenets of CDR while respecting the need for upholding data ownership rights, privacy and the right of an individual to monetise their own data (i.e. open banking).

Thus, striking a balance between the oversight of digital ethics "within" AI/FinTech-enabled FS in reducing the use of unbiased AI decision-making algorithms (where possible) and opening access to digital technologies through facilitating connectivity, skills, and tools relative to digital finance. In addition, "by" the organisational commitment to cultivating societal digital maturity (understanding), leading to empowered choice, decision-making and well-being (mental health) including the teams across the AI-lifecycle, therefore, embedding purpose and trust to drive AI governance adoption (Wade 2020; Coeckelbergh 2020). Hence, engaging in responsible recycling practices reinforces the DRC via reappropriation of digital devices and promoting the circular economy (Geng et al. 2019) through schools and marginalised areas with collaborative initiatives—the Learning Foundation's Digital Access For All (DAFA)[14] offsetting power consumption with renewables for offices and sustainable IT strategies. Finally, the CDR model prescribes appointing a Digital Ethics Council/CDR Advisory Board[15]. We argue that all companies require this aspect of CDR whether that be digital, data or AI focused, or all three. Crucially, this role will and must be complementary to, and of the code, as a dynamic iterative AI governance framework in which the DRC operates and extends responsibility for digital, data or AI services once launched into digital society. Of course, there is a caveat that CDR is not the only collaborative framework and is voluntary, a potential flaw we observed in implementing the DEAR and challenging for smaller AI/FinTech-enabled FS to achieve, but councils and boards can be shared while Dörr's (2020) text offers practical CDR implementation. CDR creates cultural change to avoid invoking "tick-box" compliance. Early adopters are the Swiss organisation, Ethos, (Fig. 4) demonstrating consideration of the CDR's tenets, and we witness similar implementation across Germany, France, the UK, the USA, China, and South Korea.[16]

---
[14] See https://digitalaccessforall.co.uk/
[15] Interchangeable term
[16] See https://corporatedigitalresponsibility.co.uk/links

1. Establish a digital responsibility code

2. Ensure transparency with stakeholders on digital practices and footprint

3. Comply with the highest standards of data processing and protection

4. Establish ethical principles for AI use

5. Exclude sensitive activities related to digitalisation

6. Ensure a fair and responsible social transition

7. Help reduce the environmental footprint of digital technology

Fig. 4 Ethos (2020) "A Method of Digital Responsibility"

Our Digital Futures

Society mistrusts AI systems, yet gradually we have succumbed to accept algorithms making potentially life-changing decisions as a daily "norm". Specifically, within financial services, mistrust is higher post-GFC, where customers remain cautious of this sector's interests combined with the power afforded by AI/FinTech-enabled FS. We have focused on "what" needs to be considered in governing AI systems and building trust—risk reduction of the potential harms that technologies inflict on our digital society, whereby the underpinning science is not fully understood. Who is responsible to ensure our digital safety as more AI systems are free to operate devoid of human oversight? Responsibility is much debated within the moral and philosophical literature; we have framed the normative aspect and complexity underpinning AI science and its associated ethical principles. Despite a plethora of ethical principles and guidance including the recent DEAR, precisely "how" organisations frame actions and governance and build a digital ethics culture over today's tendency to opt for the path of least resistance via the tick-box mentality remains challenging. One potential method of the "how" is "Corporate Digital Responsibility" combining the appropriate use of digital and data *within* and *by* the organisation impacting across the social-societal, economic, and environmental systems. Although voluntary for organisations, CDR members draw from their collaborations across standardisation networks centring on better AI-enabled systems (IEEE, iTechlaw and ForHumanity), enabling the equitable digital society to come to fruition. However, collaboration is key, while recognising humans' dark side in resisting conformity (Klotz and Neubaum 2016); the complexity between our normative and symbiotic selves will continue to be fluid and unpredictable, reflecting AI/FinTech-enabled FS

adoption. We must meet this challenge to redress the equilibrium of interests between the current perceived puppet—puppet-master space and avoid an uncontract dystopia. CDR, we posit, can potentially differentiate organisations, facilitating the gaining and maintaining of stakeholder trust and driving competitive advantage.

References

Abbott, R. (2018). *The Reasonable Computer: Disrupting the Paradigm of Tort Liability*, 86 George Washington Law Review 1.

Ackerman, J. (2004). Co-Governance f or Accountability: Beyond "Exit" and "Voice", *World Development*, 32(3): 447–463. https://doi.org/10.1016/j.worlddev.2003.06.015

Aitken, M., Toreini, E., Carmichael, P., Coopamootoo, K., Elliott, K., and van Moorsel, A. (2020), Establishing a Social License for FinTech: Reflections on the role of the private sector in pursuing ethical data-driven innovation, *Big Data and Society*, 1-15. https://doi.org/10.1177/2053951720908892

Alford, R. R. (1975), *Health Care Politics: Ideological and Interest Group Barriers to Reform,* Chicago, IL: University of Chicago Press.

Alford, R., R., & Friedland, R. (1975) Political Participation and Public Policy. *Annual Review of Sociology,* 1, 429-479.

AlgorithmWatch's Global AI Ethics Inventory (2020). Available at: https://algorithmwatch.org/en/ai-ethics-guidelines-inventory-upgrade-2020/ and full inventory: https://inventory.algorithmwatch.org/ (Accessed: 10th April 2021).

M. Arnold, Bellamy, R. K. E., Hind, M., Houde, S., Mehta, S., Mojsilovi' C A., Nair, R., Ramamurthy, N. K., Reimer, D., Olteanu, A., Piorkowski, D., Tsay, J., and Varshney, K. R. (2019). *FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity*, arXiv:1808.07261v2 **[cs.CY].**

Bell, F. (2020). *Culture, Conduct and Ethics in Banking: Principles and Practice.* London: Kogan Page.

Bloomberg. (2021, 14 April). *EU Set to Ban Surveillance, Start Fines Under New AI Rules*. Available at: https://www.bloomberg.com/news/articles/2021-04-13/eu-poised-to-set-ai-rules-that-would-ban-surveillance-scoring (Accessed: 15 April 2021).

Briggs, D.S. (2021). COVID-19: A strife of interests for us all and what problem are we attempting to solve? *Asia Pacific Journal of Health Management*, 16(1): i693. Available at: https://doi.org/10.24083/apjhm.v16il.693

Bryson, J. (2021). Two ways AI technology is like Nuclear technology. Available at: https://joanna-bryson.blogspot.com/2021/04/two-ways-ai-technology-is-like-nuclear.html?m=1 (Accessed: 18 April 2021).

Byrne, D. S. (2010). Comparison, Diversity and Complexity. IN Cilliers, P., and Preiser, R. (eds.). (2010). *Complexity, Difference and Identity,* Issues in Business Ethics, 26, Dordecht, Heidelberg: Springer: 61-75.

Carroll, A. B. (1991). The pyramid of corporate social responsibility: toward the moral management of organizational stakeholders. *Bus Horiz*, 34(4): 39–48.

Checkland, K., Harrison, S. and Coleman, A. (2009) 'Structural Interests' in Health Care: Evidence from the Contemporary National Health Service. *Journal of Social Policy,* 38(4): 607-625.

Cilliers, P. (1998). *Complexity and postmodernism: Understanding complex systems.* London: Routledge.

Cilliers, P. (2005). Complexity, deconstruction and relativism. *Theory, Culture, Society,* 22(5): 255-267.

Coeckelbergh, M. (2020). *AI Ethics,* The MIT Press Essential Knowledge Series, Cambridge: MIT Press.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Math. Control Signal Systems,* 2(4): 303–314. https://doi.org/10.1007/BF02551274

Dietz, G., and Den Hartog, D. N. (2006). Measuring trust inside organisations, *Personnel Review*, 35(5): 557–588.

Dietz, G., and Gillespie, N. (2012). *Recovery of Trust: Case Studies of Organisational Failures and Trust Repair*. Vol 5. London: Institute of Business Ethics.

Dörr, S. (2020) *Praxisleitfaden Corporate Digital Responsibility. Unternehmerische Verantwortung und Nachhaltigkeitsmanagement im Digitalzeitalter*. Springer Gabler, Berlin. (English edition in preparation Corporate Digital Responsibility. Managing Corporate Responsibility and Sustainability in the Digital Age. Springer, Berlin).

Durodié, C. (2019). *Decoding AI in Financial Services: Business implications for Boards and Professionals*, London: Blissetts.

Edelman (2019a). *Artificial Intelligence Survey*. The research was developed by the Edelman AI Centre of Expertise with input from the World Economic Forum. Available at: https://www.edelman.com/research/2019-artificial-intelligence-survey (Accessed: 12 September 2019).

Edelman (2019b). *19th Annual Trust Barometer: Financial Services*. Available at: https://www.edelman.com/research/trust-in-financial-services-2019 (Accessed: 24 June 2019).

Elton, D. C. (2020). Self-explaining AI as an alternative to interpretable AI, *arXiv*:2002.05149v5 [cs.AI].

Financial Conduct Authority (FCA) (2016). *Behaviour and Compliance in Organisations: Encouraging debate among academics, practitioners, and policymakers in all aspects of financial regulation*, Occasional Paper 24. London: FCA. Available at: https://www.fca.org.uk/publication/occasional-papers/op16-24.pdf

Floridi, L. (2019). Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical, *Philosophy & Technology*, 32:185–193. https://doi.org/10.1007/s13347-019-00354-x

General Data Protection Regulation (GDPR) (2018). Available at https://www.gov.uk/data-protection#:~:text=The%20Data%20Protection%20Act%202018%20is%20the%20UK's%20implementation%20of,used%20fairly%2C%20lawfully%20and%20transparently (Accessed: 1st June 2018).

Geng, Y., Sarkis, J., and Bleischwitz, R. (2019). How to globalize the circular economy, *Nature Comment*, 565: 153-163.

Hamdan, A., Hassanien, A. E., Razzaque, A., and Alareeni, B. (eds.) (2021). *The Fourth-Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success,* Studies in Computational Intelligence 935, Cham, Switzerland: Springer.

Haochen, G., and Polak, P. (2021). Artificial Intelligence and Financial Technology FinTech: How AI Is Being Used Under the Pandemic in 2020. IN Hamdan, A., Hassanien, A. E., Razzaque, A., and Alareeni, B. (eds.) (2021). *The Fourth-Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success,* Studies in Computational Intelligence 935, Cham, Switzerland: Springer, 169-186.

Hauer, T. (2018). Society and the Second Age of Machines: Algorithms Versus Ethics, *Society*, 55: 100–106. https://doi.org/10.1007/s12115-018-0221-6

Hauer, T. (2019). Society Caught in a Labyrinth of Algorithms: Disputes, Promises, and Limitations of the New Order of Things, *Society,* 56: 222–230. https://doi.org/10.1007/s12115-019-00358-5

Herden, C. J., Alliu, E., Cakici, A., Cormier, T., Deguelle, C., Gambhir, S., Grifths, C., Gupta, S., Kamani, S. R., Kiratli, Y-S., Kispataki, M., Lange, G., Moles de Matos, L., Tripero Moreno, L., Nunez, H. A. B., Pilla, V., Raj, B., Roe, J., Skoda, M., Song, Y., Ummadi, P. K., Edinger-Schons, L. M. (2021). "Corporate Digital Responsibility": New corporate responsibilities in the digital age, *NachhaltigkeitsManagementForum*. https://doi.org/10.1007/s00550-020-00509-x

Iivari, N., Sharma, S., and Ventä-Olkkonen, L. (2020). Digital transformation of everyday life – How COVID-19 pandemic transformed the basic education of the young generation and why information management research should care? *International Journal of Information Management*, 55: 102183. https://doi.org/10.1016/j.ijinfomgt.2020.102183

Klotz, A. C., and Neubaum, D. O. (2016). Article Commentary: Research on the Dark Side of Personality Traits in Entrepreneurship: Observations from an Organizational Behavior Perspective, *Entrepreneurship Theory and Practice,* (January): 1042-2587. https://doi.org/10.1111/etap.12214

Koshiyama, A., Kazim, E., Treleaven, P., Rai, P., et al. (2021). *Towards Algorithm Auditing A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms*. Available at SSRN: https://ssrn.com/abstract=3778998.

Kowalikova, P., Polak, P., and Rakowski, R. (2020). The Challenges of Defining the Term "Industry 4.0", *Society,* 57: 631-66. https://doi.org/10.1007/s12115-020-00555-7

Kunneman, H. (2010). Ethical Complexity. In Cilliers, P., and Preiser, R. (eds.). *Complexity, Difference and Identity,* Issues in Business Ethics, 26, Dordecht, Heidelberg: Springer: 131-163.

Lepore, J. (2020). *If then: How one data company invented the future.* John Murray: London.

Lewis, R. (2021). *The Crypto-Currency Revolution: Finance in the age of Bitcoin, blockchains and tokens,* London: Kogan Page.

Lindenberg, N. (2014). Definition of Green Finance, DIE mimeo. Available at SSRN: https://ssrn.com/abstract=2446496

Lobschat, L., Mueller, B., Eggers, F., Brandimarte, L., Diefenbachf, S., Kroschkea, M., and Wirtzh, J. (2021). Corporate digital responsibility, *Journal of Business Research,* 122: 875–888. https://doi.org/10.1016/j.jbusres.2019.10.006

Maddeo, M., and Floridi, L. (2018). How AI can be a force for good: An ethical framework will help to harness the potential of AI while keeping humans in control, *Science*, 361(6404): 751-2.

Microsoft (2018). *The Future Computed: Artificial Intelligence and its role in society*, (foreword by Smith, B and Shum, H.), Redmond: Microsoft.

Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans,* Milton Keynes: Pelican Books (Penguin Random House, UK).

Mittelstadt B. (2019). AI Ethics – Too Principled to Fail: Principles alone cannot guarantee ethical AI, *Nature Machine Intelligence,* 1: 501–507.

Molnar, C. (2019). *Interpretable Machine Learning: A Guide for making Black-Box models explainable.* https://christophm.github.io/inter\pretable-ml-book.

Morley, J., Floridi, L., Kinsey, L., and Elhalal, A. (2020). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices, *Sci Eng Ethics,* 26: 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mokander, J., and Floridi, L. (2021). Ethics as a service: a pragmatic operationalisation of AI Ethics. Available at: SSRN: https://ssrn.com/abstract=3784238 or https://doi.org/10.2139/ssrn.3784238

North, N. and Peckham, S. (2001). Analysing Structural Interests in Primary Care Groups. *Social Policy and Administration,* 35(4): 426-440.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information,* Harvard University Press.

Pasquale, F. (2017*). Secret Algorithms Threaten the Rule of Law*, MIT Technology Review.

Pasquale, F. (2020). *New Laws of Robotics: Defending Human Expertise in the Age of AI*, Cambridge: The Belnap Press of Harvard University Press.

Pedersen, N. (2021). *Financial Technology: Case Studies in FinTech Innovation,* London: Kogan Page.

Perez, C. C. (2019). *Invisible Women: Exposing Data Bias in a World Designed for Men,* London: Vintage (Penguin Random House, UK).

Roitblat, H. L. (2020). *Algorithms Are Not Enough: Creating General Artificial Intelligence,* Cambridge: The MIT Press.

Shaw, P. (2020). Building Trust through Sound Governance. In Chishti, S., Bartoletti, I., Leslie, A., and Millie, S. M. (eds.), *The AI Book: The Artificial Intelligence Handbook for Investors, Entrepreneurs and FinTech Visionaries,* Chichester: John Wiley & Sons Ltd, 175-179.

Shaw, P. (2021). "Context Matters": The Law, Ethics and AI. In Hervey, M., and Lavy, M. *The Law of Artificial Intelligence,* Croydon: Sweet & Maxwell (Thomas Reuters), CPI Group, UK, 31-66.

Taghizadeh-Hesarya F., and Yoshinob, N. (2019). The way to induce private participation in green finance and investment, *Financial Research Letters,* 31: 98-103. https://doi.org/10.1016/j.frl.2019.04.016

The Verge. (2021). *The EU is considering a ban on AI for mass surveillance and social credit scores.* Available at: https://www.theverge.com/2021/4/14/22383301/eu-ai-regulation-draft-leak-surveillance-social-credit (Accessed: 15 April 2021).

Toreini, E., Aitken, M., van Moorsel, A., Elliott, K., and Coopamootoo, K. (2020), The relationship between trust in AI and trustworthy machine learning technologies, *ACM Fairness, Accountability and Transparency Conference 2020*, Barcelona, Publisher: ACM. https://doi.org/10.1145/3351095.3372834

Ülgen, F. (2017). Financialization and Vested Interests: Self-Regulation vs. Financial Stability as a Public Good.*Journal of Economic Issues*, 51:2, 332-340, https://doi.org/10.1080/00213624.2017.1320512

van Dijk, J. (2019). *The Digital Divide.* Cambridge: Polity.

van Vuuren, H. J. (2020). The Disclosure of Corporate Governance: A Tick-Box Exercise or Not? *International Journal of Business and Management Studies,* 12(1): 50-65.

Wade, M. (2020). Corporate Responsibility in the Digital Era, MIT Sloan Management Review (28 April). Available at: https://sloanreview.mit.edu/article/corporate-responsibility-in-the-digital-era/ (Accessed: 1 May 2020).

Wiederhold, B. K. (2020). Connecting Through Technology During the Coronavirus Disease 2019 Pandemic: Avoiding "Zoom Fatigue", *Cyberpsychology, Behavior, and Social Networking*, 23(7): 437-440.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for the Future of the New Frontier of Power,* London: Profile Books Ltd.

Karen Elliott is an associate professor in Enterprise/Innovation, specialising in sociotechnical interdisciplinary research between business, technology, and social sciences. Named "Standout #35 Women in FinTech Powerlist by Innovate Finance", she is the co-investigator of FinTrust (EPSRC) and Finclusion (Gates Foundation/Turing Institute) projects with Prof van Moorsel and forms part of the IEEE Ethical AI and ForHumanity Committees. Her work examines FinTech, trust, digital ethics, and Corporate Digital Responsibility (CDR), to promote an equitable digital society.

Rob Price is a director at Alchemmy, a UK-based management consultancy, and previously COO/Deputy CEO Worldline, and CDO at Atos, UK. At Worldline, he co-founded the principle of Corporate Digital Responsibility, where he led a global research workstream in a global thinktank led by Thierry Breton. As part of the MCA's Year of Digital, he defined "What is Digital?" in 2015, and in 2016 was listed in Onalytica's Top 50 most influential on Digital Transformation.

Patricia Shaw is CEO of Beyond Reach Consulting Ltd, providing AI/data ethics strategy, public policy engagement, bespoke AI/data ethics governance, expert advice, and advisory board services, across financial services, public sector (Health and EdTech), and smart cities. A non-practising solicitor, public speaker, author, and listed on 2021 100 Brilliant Women in AI Ethics™, including Society for Computers and Law, chair and trustee; iTechlaw, vice-chair of the AI committee; and IEEE's Ethical Certification Program for Autonomous Intelligent Systems, and as ForHumanity's Independent Audit of AI Systems Fellow.

Tasos Spiliotopoulos is a Human-Computer Interaction researcher with a focus on social technologies and human-centred design. Tasos is a postdoctoral research associate at the Newcastle University Business School in the UK, where he currently researches trust in the financial industry. Key topics in his research include privacy, trust, cybersecurity, computational social science, machine learning, digital identities, and computer-mediated communication.

Magdalene Ng is a research associate on the FinTrust project on the human aspects of trust and privacy in FinTech, with a background in Psychology. She also spends her time as an innovation fellow at the Centre for Digital Citizens, where she is interested in investigating the intersections of human behaviour, technology and crime.

Kovila Coopamootoo , PhD, is a research fellow in User-Centred Privacy & Security at the Newcastle University. Her research involves a deep understanding of users and their privacy and security practices, given their specific needs and vulnerabilities, and in developing social, collaborative practices for protection and safety online.

Aad van Moorsel is a professor in Computer Science specialising in cyber security, specifically interdisciplinary research at the intersection of computing, psychology, and social sciences. He has been principal investigator of several interdisciplinary projects, funded by UK, EU, and USA funding bodies. His work includes the study of economic mechanism design to incentivise healthy security behaviour, the research in choice architecture for that purpose and more recently the study of trustworthy AI systems.