JESUS COLLEGE
CAMBRIDGE
UK - China Global Issues Dialogue Centre

Discussion Paper

Data Ethics and Multinational Technology Companies

Sam Gilbert & Peter J. Williamson

Introduction

It is difficult to overstate the impact of digital technology. This has been rapid and wide-ranging, affecting everyday life, markets, politics, and society.

Much human life is now lived and experienced digitally. More than 40% of the world's population – some 3.2 billion individuals[1] – owns a smartphone, with a further 800 million owning feature phones[2]. Tencent's WeChat has 1.1 billion users concentrated in the Greater China area, over 80% of whom use the app for more than 10 minutes every day[3]. Facebook is even more deeply penetrated into daily life: between them, the Facebook "blue app", WhatsApp, Messenger, and Instagram are used for an average of almost an hour a day[4] by more than 2.4 billion people[5]. Digital technology has made it significantly faster and easier to start, grow, and geographically expand new businesses, with e-commerce now accounting for more than 15% of retail sales in the United States[6]. Nor is digital technology confined to the developed world: there are many more internet users in India (560m) than in the USA (312m); many more in Nigeria (111m) than in Germany (79m)[7].

As a result, the technology companies which provide digital devices and software applications have become the world's most valuable corporations. In 2008, the Top 10 list of the world's largest companies by market capitalisation contained only one technology company, Microsoft. By the end of 2018,

[1] Statista (2019a), "Number of Smartphone users worldwide", https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/, accessed 30 December 2019
[2] Statista (2019b), "Mobile Interet Usage Worldwide", https://www.statista.com/topics/779/mobile-internet/, , accessed 30th December 2019
[3] Statista (2019c), "Most popular global messenger apps", https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/, accessed 30th December 2019
[4] Meeker, M. (2018), "Internet Trends", https://www.kleinerperkins.com/perspectives/internet-trends-report-2018/, accessed 3 January 2020: 114
[5] Facebook (2019), "Q3 2019 Earnings Presentation", https://s21.q4cdn.com/399680738/files/doc_financials/2019/q3/Q3-2019-Earnings-Presentation.pdf, accessed 3 January 2020
[6] Meeker, M. (2019), "Internet Trends 2019", Bond Capital, https://www.bondcap.com/report/itr19/#view/title: 20
[7] Internet World Stats (2019), https://www.internetworldstats.com/list2.htm, accessed 3 January 2020

companies in oil, retail, telecoms, banking, and fast-moving consumer goods had been supplanted by Alibaba, Amazon, Apple, Facebook, Tencent, and Google's parent Alphabet[8].

The power of big technology companies is different in kind (not merely in degree) to the power of previous generations of leading multinationals. It is a kind of power which extends far beyond economic influence. For example, it extends into the geopolitics of the Eastern Congo, where the coltan that is integral to all digital devices is mined[9]. It extends into national security, where it has been co-opted by governments to enable covert state monitoring of citizens' private communications, as Edward Snowden revealed[10]. It extends into democratic politics, where it has been exploited by both legitimate and disingenuous actors seeking to influence the outcome of referendums and elections[11]. It is clearly a disruptive social force, as well as an enabler of economic growth and human progress.

It is within this context of digital disruption that this report seeks to make its contribution. The report is intended primarily for a practitioner audience: that is, for executives and employees at technology companies, for policymakers, for lawyers, and for tech entrepreneurs. Its focus is *data ethics*. The term "data ethics" refers to the latest evolutionary stage of technology ethics. From the 1970s to the 2000s, ethical questions relating to technology tended to be discussed either at the level of computers, or at the level of computers' human designers and users. The digital era necessitates engaging with ethics at a more granular level – the level of data[12]. Other scholars use the term "digital ethics": for the practical purposes of this report, "digital ethics" and "data ethics" are synonymous. Of course, both overlapping and distinct ethical issues arise in relation to other forms of technology, such as nanotechnology and biotechnology, but these are not within our scope.

> **Data Ethics** is "a new branch of ethics that studies and evaluates the moral problems related to data (including generation, recording, curation, processing, dissemination, sharing and use), algorithms (including artificial intelligence, artificial agents, machine learning and robots) and corresponding practices (including responsible innovation, programming, hacking and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values)."
>
> – Luciano Floridi & Mariarosaria Taddeo

The report divides into three parts. In **Part One**, we provide an overview of the data ethics field, drawing out the key themes in contemporary academic work: privacy and surveillance; bias, discrimination, and injustice in algorithmic decisioning; encoding of ethical assumptions in autonomous vehicle systems; artificial general intelligence as an existential risk to humanity; software user interface design as an impediment to human flourishing; job displacement from machine-learning and robotics; and monetary compensation for personal data use. Using data from Google Trends, we assess the extent to

[8] Johnston, S. (2019) "Largest companies 2008 vs. 2018, a lot has changed", https://milfordasset.com/insights/largest-companies-2008-vs-2018-lot-changed; 3rd January 2020
[9] Smith, J. H. (2011), "Tantalus in the Digital Age: Coltan ore, temporal dispossession, and 'movement' in the Eastern Democratic Republic of the Congo." *American Ethnologist* 38, no. 1: 17-35
[10] Greenwald, G. and MacAskill, E. (2013), "NSA Prism program taps in to user data of Apple, Google and others", *The Guardian*, https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data, accessed 3rd January 2020
[11] For example: Office of the Director of National Intelligence (2017), "Assessing Russian Activities and Intentions in Recent US Elections", https://www.dni.gov/files/documents/ICA_2017_01.pdf, accessed 3rd January 2020
[12] Floridi, L. & Taddeo, M., (2016). "What is data ethics?" *Philosophical transactions*. Series A, Mathematical, physical, and engineering sciences, 374(2083), 28 December 2016, Vol.374(2083).

which these themes reflect the concerns of the general public. We then outline current responses to data ethics issues by governments, intergovernmental organizations, and technology companies, highlighting gaps and opportunities.

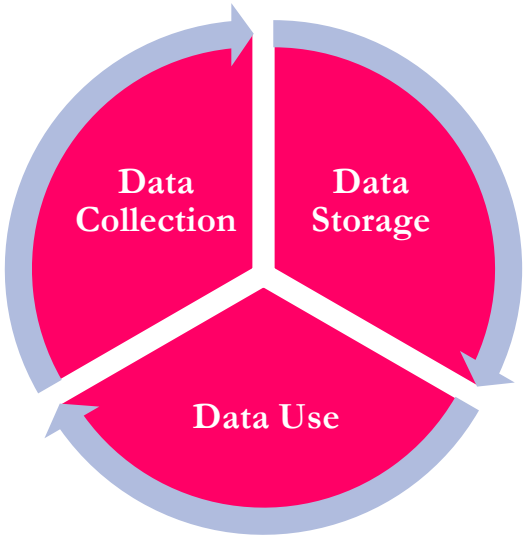
In **Part Two**, we consider how technology companies should approach data ethics in practice. Briefly reviewing the commercial impact of past environmental disasters on firms in the oil and petrochemical industries, we demonstrate that engaging with data ethics is a practical necessity for all multinational technology companies. But how are such companies to reconcile variations in ethical traditions and norms across the different territories in which they operate? Using a case study of Shanghai Rodway, the sometime China subsidiary of the global information services firm Dun & Bradstreet, we show that an "overlapping consensus" between different ethical traditions may be achievable on many questions of data ethics. Turning to philosophical theories of ethics, we identify *virtue ethics* as the most suitable framework for technology companies' ethical decision-making, suggesting that the best question to ask is "What kind of company should we be?" This is illustrated with a case study of the web infrastructure company Cloudflare, and a worked example of a telecommunications provider diversifying into value-added services.

Part Three offers tangible recommendations. We summarize current best practice on data ethics in technology companies, before sketching some fresh ideas for new actions which companies might take. Many existing actions – such as the establishment of AI Ethics Boards and guidelines for ethical decision-making – are focused on mitigating risks associated with the collection, storage, and use of data; we therefore focus on ways that technology companies can act to maximize the good they do in the world. We propose a set of initiatives involving the opening up of data held by technology companies as a source of public good, and a process of co-creation between workers at high risk of technological displacement, technologists, and AIs to design the jobs of the future.

1. The Data Ethics Landscape

1.1 Why data ethics?

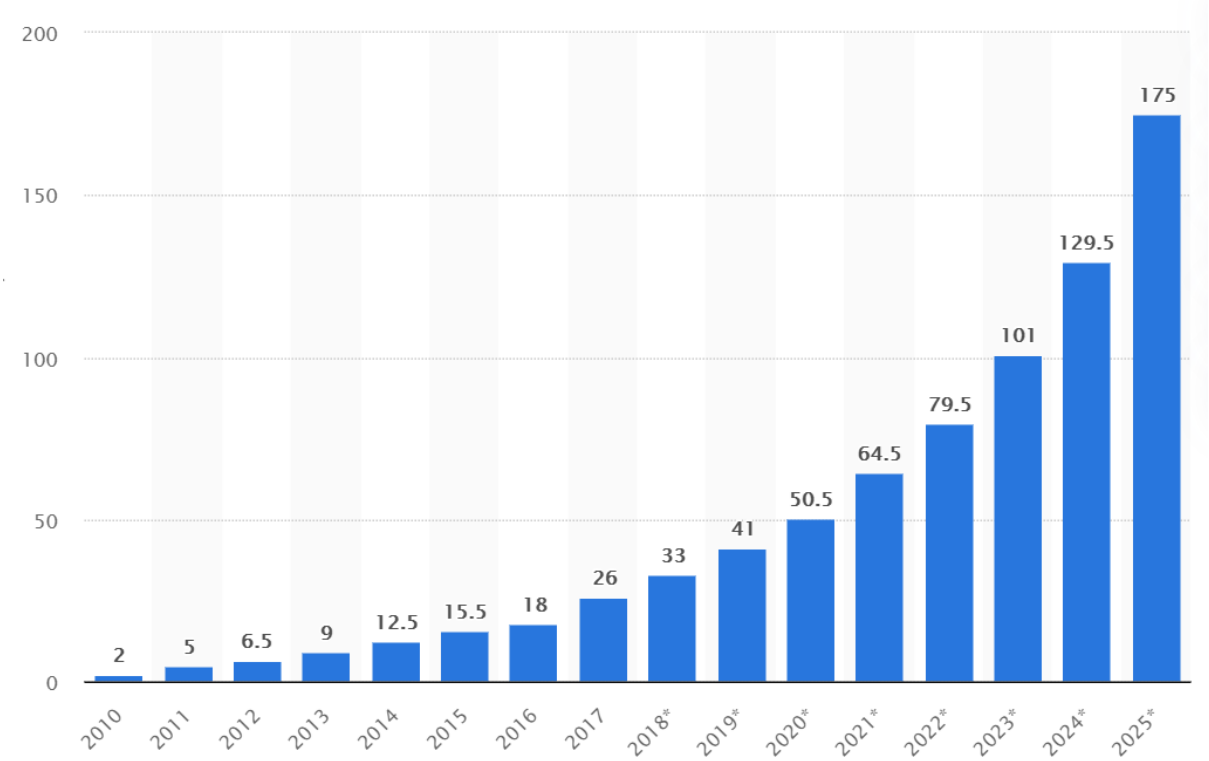
The collection, storage, and use of data about individuals – or "personal data" – is integral to technology companies' business models and to the wider digital ecosystem in which they operate. Ethical questions arise in relation to personal data for three reasons. Firstly, there are circumstances in which collecting and storing personal data may infringe individuals' human rights. Secondly, in liberal societies, personal data is conceptualized as a form of private property, to which property rights therefore apply. Thirdly, and most importantly, the use of personal data may harm individuals.



Historically, potential harms related to the use of data about race, ethnicity, and/or sexuality by the state to facilitate coercion of individuals or groups. In the era of information technology, the collection and storage of personal data by corporations for commercial purposes created the potential for other harms, including identity fraud.

The ubiquity of digital technology amplifies this potential on multiple dimensions. *New types of data* are being produced by individuals' profiles on social networks, their online activity, their use of mobile devices, digital recording of their physical characteristics, and by internet-connected sensors in the physical world – respectively "profile data", "behavioural data", "location data", "biometric data", and "IoT data". The overall *volume of data* being produced has increased exponentially (see Figure 1). Combined with advances in machine-learning, this greater variety and volume of data is enabling an increasing range of decisions affecting individuals to be made *algorithmically*, in both the private and public sectors. Meanwhile, *new business models* which derive monetary value from personal data have emerged, including highly targeted digital advertising.

Figure 1: "Data Proliferation": volume of data created worldwide from 2010 to 2025 in zetabytes[13]



Furthermore, the storage of these new types of data by a wide range of organizations in and of itself creates new risks of misappropriation by malevolent actors. The plausibility of phishing attacks and the leverage of ransomware (extortion) attacks may be increased by access to more personal data. Sensitive personal data may be used to inflict various forms of social cruelty, including doxxing, trolling, public shaming, and cyberbullying. In conjunction with targeting technologies such as Facebook Lookalike Audiences, personal data may be used for illegal forms of profiling, or in ways that are widely regarded as illegitimate (for example, in psychographic targeting models or voter-suppression campaigns)[14].

New forms of harm: examples

Doxxing: Maliciously publishing personal information about a specific individual on the internet

Trolling: Posting unsolicited comments online with the intention of causing hurt or provoking an emotional reaction from an individual or group

Cyberbullying; The use of digital messaging applications to send threatening or insulting messages to an individual

1.2 Key concerns of academic literature

Accordingly, academic literature increasingly considers questions of technology ethics at the level of data. For this report, we compiled a

[13] Statista (2019), https://www.statista.com/statistics/871513/worldwide-data-created/, accessed 3rd January 2020
[14] Propbulica (2017), "Facebook (Still) Letting Housing Advertisers Exclude Users by Race", https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin, accessed 30 December 2019

database of 953 peer-reviewed journal articles on data ethics topics, the titles of which are visualized in Figure 2.

Figure 2: Word cloud visualization of all article titles in the authors' database of academic literature on digital ethics (n=953). Frequencies are shown in parentheses. Visualization: Tagcrowd.

From analysis of the database, we draw out seven key themes for discussion:

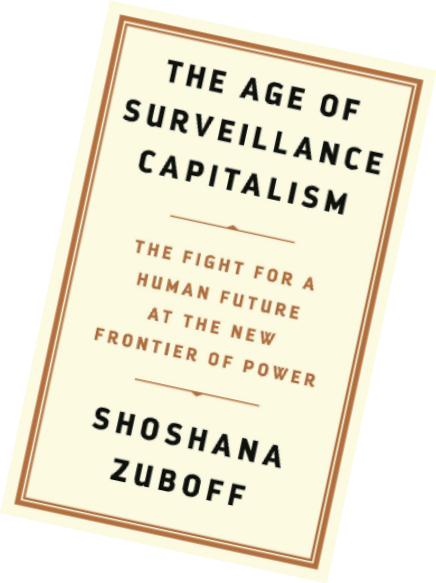
- Privacy and surveillance
- Bias, discrimination, and injustice in algorithmic decisioning
- Encoding of ethical assumptions in autonomous vehicle systems
- Artificial general intelligence as an existential risk to humanity
- Software user interface design as an impediment to human flourishing
- Job displacement from machine-learning and robotics
- Monetary compensation for personal data use

Privacy and surveillance

The near-ubiquitous collection and storage of personal data in technologically-advanced societies is widely regarded by scholars as a threat to the human right to privacy[15]. The continuous circulation of behavioural data, such as cookie data, is inherent to the functionality of social networking apps, e-commerce, online banking and many other internet services. Ambient computing devices such as Amazon Echo and Google Home (colloquially known as "smart speakers") are designed to be used in intimate spaces, and may therefore be seen as violating the private sphere of the home. Still more controversial are internet-connected toys such as Hello

[15] Global Internet Liberty Campaign (2020), "PRIVACY AND HUMAN RIGHTS: An International Survey of Privacy Laws and Practice", http://gilc.org/privacy/survey/intro.html, accessed 3 January 2020

Barbie which may impinge on the emotional privacy of children[16]. The popularity of wearable devices such as Apple Watch and Fitbit (recently acquired by Google[17]), pre-installation of wellbeing apps such as Samsung Health on smartphones, and the deployment of facial recognition technology all mean that increasing volumes of biometric data are held in databases controlled by technology companies, compromising bodily privacy[18]. A corollary of this in the Global South are the biometric databases underpinning both state-run citizen registration programmes and the fingerprint and retinal scanning stations which form an integral part of Europe's counter-migration technology infrastructure[19]. Finally, the growth of "smart cities" is expected to further widen the scope of personal data collection and storage, as citizens' physical movement through the built environment and interactions with public services ranging from transportation to waste disposal become trackable through sensors, RFID tags, and contactless payments.



There is increasing support for the view that many such data collection practices are covert and non-consensual, and therefore amount to illegitimate surveillance. Scholars such as the Harvard Business School professor Shoshana Zuboff, who coined the influential phrase "Surveillance Capitalism"[20], argue that users have not, in any meaningful sense, consented to the use of their profile and behavioural data in the targeted digital advertising which forms the commercial basis of many technology companies (notably Google and Facebook)[21]. While it is acknowledged that such uses may be allowed for in the companies' Terms of Service and/or Data Use Policies, the assertion is that the length and opacity of these documents are such that any consent conferred by them cannot be regarded as *informed* consent. Moreover, it is claimed that it is practically impossible for users to opt out of their profile and behavioural data being used without losing access to the services Google and Facebook provide. Turning to the use to which personal data is put, Surveillance Capitalism theorists argue that highly targeted digital advertising constitutes manipulation, and should therefore be seen as an unethical exercise of coercive power.

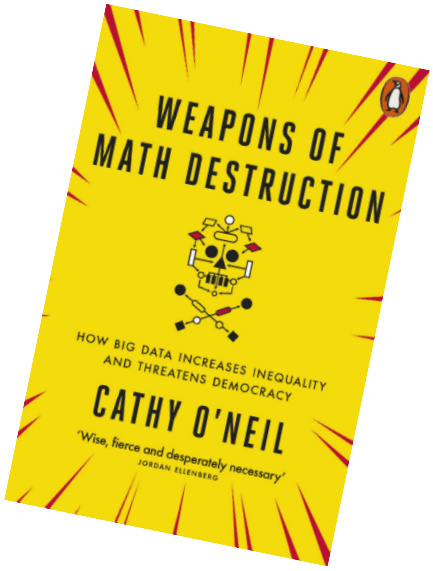
Bias, discrimination, and injustice in algorithmic decisioning

Algorithmic decision-making is not a new phenomenon. Since the 1980s, financial services companies have used automated scorecards to decide whether to accept or reject applications for loans, credit cards, car finance plans, and mortgages. Drawing on data about individuals held in credit bureaux, scorecards had supplanted human bank managers as the chief decision makers on

[16] Rosner, G. and Kenneally, E. (2018), *Clearly Opaque: Privacy Risks of the Internet of Things*, IoT Privacy Forum
[17] Gartenberg, C. (2019), "Google buys Fitbit for $2.1bn", *The Verge,* https://www.theverge.com/2019/11/1/20943318/google-fitbit-acquisition-fitness-tracker-announcement, accessed 3rd January 2020
[18] By the end of 2018 there were 52m users of wearables in the USA, up ~100% on 2014. Wurmser, Y. (2019), "Wearables 2019", eMarketer, https://www.emarketer.com/content/wearables-2019
[19] Meaney, T. (2019), "Who's your dance partner?", *London Review of Books* Vol. 41 No. 21: 35-39
[20] Zuboff, S., (2015). "Big other: surveillance capitalism and the prospects of an information civilization". *Journal of Information Technology*, 30(1), pp.75–89
[21] Zuboff, S., (2019). *The age of surveillance capitalism: the fight for the future at the new frontier of power*, London: Profile Books. Kindle Edition.

the extension of credit by the turn of the century. Similar models were adopted by direct-to-consumer insurance brands to automate underwriting decisions.

Prior to the 2008 subprime mortgage crisis, which highlighted widespread irresponsible consumer lending, there is little evidence that the use of algorithmic decision-making in financial services was regarded as an ethical issue. However, both data proliferation and advances in machine learning during the digital era have enabled algorithmic decision-making to be extended to other domains, including criminal justice and policing – often facilitated by technology companies such as Palantir. It is this which has drawn the attention of scholars. In courtrooms, recidivism algorithms are used in sentencing and parole decisions[22]. "Predictive policing" models are used to determine where law enforcement personnel should be deployed in anticipation of potential crimes[23]. Scholars have demonstrated that both the unconscious biases of algorithm designers and their models' reliance on historic data can reproduce racial, ethnic, and gender discrimination. It is widely argued that the first step to mitigating this risk is for organizations to accept that they have a moral responsibility to make their algorithms both transparent to regulators and explainable to the individuals who are affected by them[24].



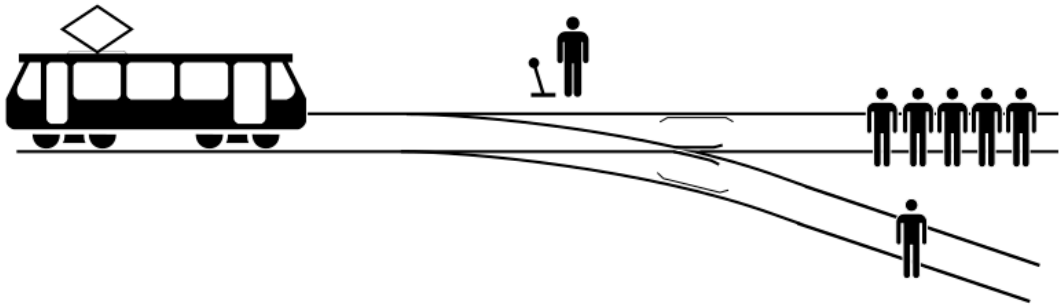
Encoding of ethical assumptions in autonomous vehicle systems

Road traffic accidents are a major cause of death globally, accounting for 1.4m fatalities in 2016[25]. While developers of self-driving vehicles claim they will have a net favourable impact on road safety by reducing human error, their adoption by consumers and in the haulage and couriering sectors will inevitably lead to their involvement in fatal accidents. Indeed, five deaths have already been caused by Tesla's automated driving system, and one by Uber's[26]. The main ethical question raised relates to the software used to determine how a self-driving vehicle should react in the event of a probable collision. Programmers of this software are presented with a version of the ethical dilemma famously framed by the philosopher Philippa Foot as "The Trolley Problem"[27], as illustrated in Figure 3.

[22] O'Neil, C., (2017). *Weapons of math destruction: how big data increases inequality and threatens democracy*, London: Penguin Books
[23] Fry, H., (2018). *Hello world: how to be human in the age of the machine*, London: Transworld Digital
[24] See for example Gillis, T. and Simons, J. (forthcoming): "Explanation < Justification: GDPR and the Perils of Privacy"
[25] World Health Organization (2018), "The top 10 causes of death", https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death, accessed 8 January 2020
[26] Wikipedia (2020), "List of self-driving car fatalities", https://en.wikipedia.org/wiki/List_of_self-driving_car_fatalities, accessed 8 January 2020
[27] Foot, P. (1978) "The Problem of Abortion and the Doctrine of the Double Effect" in *Virtues and Vices* Oxford: Basil Blackwell

Figure 3: The Trolley Problem[28]

A train is out of control and heading towards five people, who are unable to move and will be killed. Should you pull the lever to divert the train, killing one person instead?



The Trolley Problem captures the divergence of views between the *deontological* and *utilitarian* schools of philosophical ethics (discussed in more detail in Part Two). Of necessity, one of these views must be encoded into autonomous vehicle software. In addition, the software needs to know how to prioritize the life of the passenger against the lives of pedestrians, bystanders, and passengers in other vehicles[29]. With many human lives potentially at stake, scholars have argued that this should not be determined by programmers, while also highlighting the challenges inherent in securing democratic consent for regulatory approaches, and cross-cultural variations in ethical preferences[30].

Artificial general intelligence as an existential risk to humanity

The view of a large sample of AI experts, surveyed by researchers in 2017, is that there is a 50% chance of AI outperforming humans in all tasks by 2060[31]. Although there is doubt and disagreement about the velocity of this risk, its severity is more certain: human intelligence is constrained by the information processing limits of biological tissue, but machines are not subject to the same limits. In other words, the potential of artificial intelligence is to all intents and purposes unconstrained. According to Nick Bostrom and other scholars at the University of Oxford's Future of Humanity Institute, AI therefore poses an existential threat to humanity[32].

This threat can be illustrated with a simple thought experiment. Imagine a machine superintelligence given the objective of calculating pi. Optimising towards its goal, it could be expected to channel all the world's resources into building a planet-sized supercomputer. Since humanity serves no useful computational purpose, we could expect it to be instrumentalized or wiped out entirely[33]. Bostrom therefore stresses the importance of ensuring that the objectives of AI systems are both aligned with human values, and rigorously specified so as to pre-empt

[28] Image from Wikimedia Commons: https://commons.wikimedia.org/wiki/File:Trolley_Problem.svg
[29] For a discussion, see Fry (2018).
[30] Awad, E., Dsouza, S., Kim, R. et al. (2018). "The Moral Machine experiment" *Nature* 563, 59–64 doi:10.1038/s41586-018-0637-6
[31] Grace, K. et al. (2017), "When Will AI Exceed Human Performance? Evidence from AI Experts", https://arxiv.org/abs/1705.08807
[32] Bostrom, N. (2014). *Superintelligence: paths, dangers, strategies*. Oxford, United Kingdom: Oxford University Press
[33] Susskind, J. (2018). *Future politics: living together in a world transformed by tech*, Oxford, United Kingdom: Oxford University Press

unintended consequences. This in turn raises further ethical questions: which values should be embedded, and who should determine them?

Software user interface design as an impediment to human flourishing

Scholars and cultural commentators have argued that the addictive nature of devices and apps is a constraint on human freedom and an impediment to human flourishing[34]. The average American now spends a total of 6.3 hours per day using digital media[35], touching their smartphone over 2,500 times[36]. Globally, 1 billon hours of Youtube videos are watched every day[37]. More than 2 billion people play interactive online games such as Fortnite[38]. Studies have suggested that heavy use of digital technology may be a factor in increases in self-harm, eating disorders, and suicidal ideation among young people[39]. It may also be a driver of more generalized feelings of distraction, dissatisfaction, boredom, frustration, and malaise[40].

Technology companies are accused of behaving unethically by using the talents of the designers, data scientists and programmers they employ to maximize "engagement", as quantified by the duration of users' sessions and the volume of their likes, comments, shares. Put simply, the claim is that digital technology is an obstacle to the Good Life, and that technology companies have prioritized their own growth and profitability over their users' wellbeing.

Job displacement from machine-learning and robotics

In 2013, research by Carl Benedikt Frey and Michael Osborne suggested that 47% of American jobs were at high risk from computerization[41]. Their paper precipitated a wave of reports and studies on the future of work, receiving over 5,000 academic citations[42]. Longer term prospects for jobs appear even more bleak: the survey of AI experts referenced above suggests there is a 50% chance that machines will have displaced all human jobs by 2140[43]. So central is work to the contemporary sense of human identity, that job displacement is an existential – as well as an economic – question[44]. It has therefore been suggested that the technology companies which prosper from automation have moral duties towards the workers whose livelihoods it puts at risk.

---

[34] Williams, J., (2018), *Stand out of our light: freedom and resistance in the attention economy*, Cambridge: Cambridge University Press
[35] Meeker (2019): 41
[36] Winnick, M. (2016), "Putting a Finger on Our Phone Obsession", Dscout, https://blog.dscout.com/mobile-touches, accessed 3rd January 2020
[37] Youtube (2020), "Youtube by the numbers", https://www.youtube.com/about/press/, accessed 3 January 2020
[38] Meeker (2019): 89ff
[39] For example: Moreno et al., (2016), "Secret Society 123: Understanding the Language of Self-Harm on Instagram". *Journal of Adolescent Health*, 58(1): 78–84
[40] Shakya, H.B. & Christakis, N.A., 2017. "Association of Facebook Use With Compromised Well-Being: A Longitudinal Study." *American Journal of Epidemiology*, 185(3), pp.203–211
[41] Frey, C. and Osborne, M. (2013), "The future of employment: How susceptible are jobs to computerisation?" *Technological forecasting and social change* 114, 254-280
[42] Google Scholar, https://scholar.google.com/citations?user=yjqqB5AAAAAJ&hl=en#d=gs_md_cita-d&u=%2Fcitations%3Fview_op%3Dview_citation%26hl%3Den%26user%3DyjqqB5AAAAAJ%26citation_for_view%3DyjqqB5AAAAAJ%3Au5HHmVD_uO8C%26tzom%3D-60, accessed 8 January 2020
[43] Grace et al (2017)
[44] Cohen, J. (2018). *Not working: why we have to stop*, London: Granta

However, acceptance that AI will lead to net job losses is not universal: although the World Economic Forum expects 75 million jobs to be lost to automation by 2022, it predicts this will be more than offset by the creation of 133 million new jobs[45]. Meanwhile, other scholars have argued that there is no essential difference between the contemporary displacement of workers by AI and robotics, and previous instances of automation-driven job displacement dating back to the first industrial revolution[46]. If either of these positions is correct, it is difficult to see job displacement as a matter of ethics (unless the legitimacy of capitalism is itself being disputed).

Monetary compensation for personal data use

Like algorithmic decisioning, monetisation of personal data by technology companies is not new. Since the 1980s, data on individuals' contact details and consumer preferences has been collected by information services firms such as Experian, Acxiom, and Claritas. This data has been sold to clients for marketing and sales purposes at a price of ~$1 per record, and used in segmentation systems and software such as Mosaic and Prism. However, in the digital era, the remarkable growth and profitability of technology companies with data-driven business models (notably Facebook and Google) has shaped a view that the economic value of personal data is now much greater than previously. Data has been called "the world's most valuable resource" by *The Economist*[47] and frequently compared by scholars to precious commodities such as oil, coal, and treasure[48].

Because personal data is treated as private property in liberal societies, it has been argued that individuals should receive monetary compensation from companies which profit from the use of their data as a matter of justice – notably by the computer science and philosophy writer Jaron Lanier[49]. Similarly, Zuboff claims that the use of personal data as a "free raw material" means that digital advertising-based business models are expropriative and extractive[50]. In this view, profile data and biometric data are forms of digital property, while behavioural data, location data, and IoT data are the surplus from a form of digital labour. Since technology companies take

[45] World Economic Forum (2018), "The Future of Jobs 2018", http://reports.weforum.org/future-of-jobs-2018/, accessed 8 January 2020
[46] Coyle, D. (2015), *GDP: A Brief but Affectionate History - Revised and expanded Edition*, Princeton University Press
[47] Economist (2017), "The world's most valuable resource is no longer oil, but data", https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data
[48] Tufekci, Z. (2017), *Twitter and tear gas: the power and fragility of networked protest*, New Haven: Yale University Press; Naughton, J. (2018), "Platform Power and Responsibility in the Attention Economy" in Moore, M. & Tambini, D. (eds.) *Digital dominance: the power of Google, Amazon, Facebook, and Apple*, New York: Oxford University Press; Susskind, J., (2018). *Future politics: living together in a world transformed by tech*, Oxford, United Kingdom: Oxford University Press
[49] Lanier, J., (2013), Who owns the future?, London: Allen Lane. See also i.am, w. (2019), "We need to own our data as a human right—and be compensated for it", *The Economist*, https://www.economist.com/open-future/2019/01/21/we-need-to-own-our-data-as-a-human-right-and-be-compensated-for-it
[50] Ibid.: Loc 71ff. See also Turner, F., (2018), "The arts at Facebook: An aesthetic infrastructure for surveillance capitalism", *Poetics*, 67, pp.53–62: 61; Turner, F. (2019) "Machine Politics: The Rise of the Internet and a New Age of Authoritarianism," *Harper's Magazine*, January, 2019, 25-33: 32

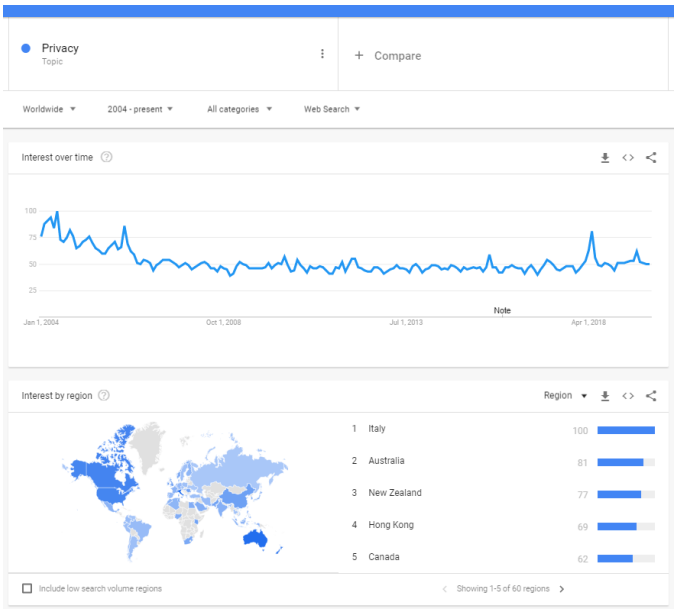
this property from their users, the condition of users is akin to serfdom[51]. The same argument for monetary compensation has been made in relation to the use of data unwittingly provided by individuals in the training of machine-learning algorithms[52].

1.3 Key concerns of digital technology users

Analysis of anonymised Google search data reveals the extent to which the concerns of academic literature relating to data ethics are shared by the general public, and also how these concerns vary by region[53].

Worldwide interest in the issue of privacy declined between 2004 and 2008, and has been relatively flat since – a spike at the time of the Cambridge Analytica scandal notwithstanding. Privacy also appears to be predominantly of interest to users in the West (See Figure 4). Indeed, a survey by Salesforce in 2018 found that 91% of American consumers are concerned about the surveillance potential of emerging technologies[54].

Figure 4: Worldwide interest in the topic "Privacy", 2004-19[55]

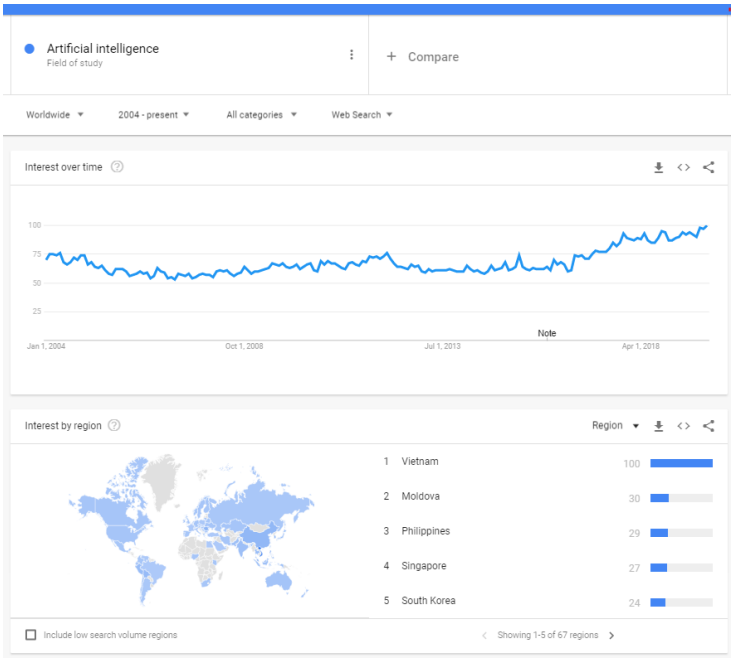


There is evidence that interest in digital user interface design has a similar, Western-oriented geographic profile. Searches for The Centre for Humane Technology – the NGO founded in 2013 by ex-Google design ethicists to campaign against the "digital attention crisis" – are predominantly in Northern Europe and North America[56]. By contrast, public interest in AI as

[51] See for example Cobbe's concept of "Produsumers". Cobbe, J. (2018a), *Big Data, Surveillance, and the Digital Citizen*. Available at SSRN: https://ssrn.com/abstract=3234984 or http://dx.doi.org/10.2139/ssrn.3234984: 101ff
[52] Posner, E. and Weyl, G. (2019), *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*, Princeton University Press
[53] For a methodological discussion of search data in social science research, see [insert link to Bennett Inst blog]
[54] Salesforce (2019), https://www.salesforce.com/company/ethical-and-humane-use/, accessed 7 January 2020
[55] Google Trends, https://trends.google.co.uk/trends/explore?date=all&q=%2Fm%2F01y4q_
[56] Google Trends, https://trends.google.co.uk/trends/explore?date=2013-01-01%202020-01-07&q=%2Fg%2F11c328bdn2

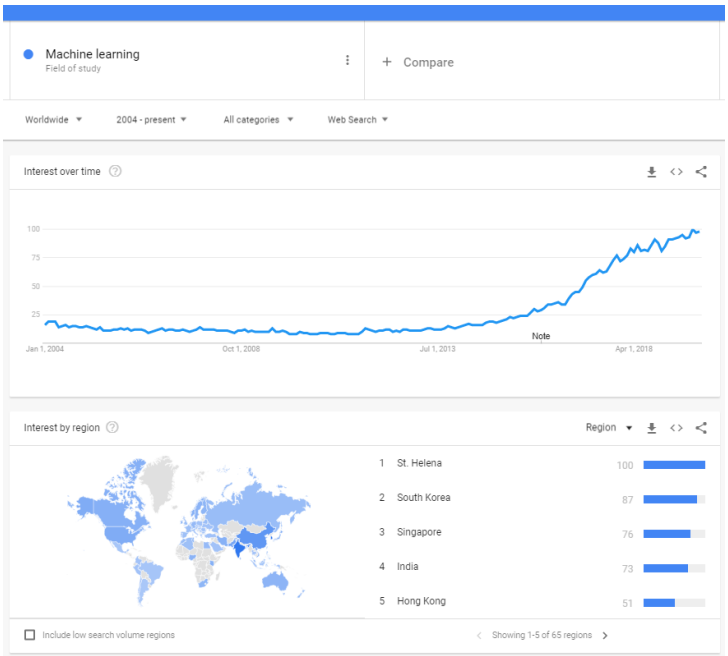
measured by Google search data has increased since 2015, and is less geographically concentrated (Figure 5).

Figure 5: Worldwide interest in the topic "Artificial Intelligence", 2004-19[57]



Within the category of AI, increases in consumer interest in machine learning have been especially pronounced, notably in Asia (Figure 6).

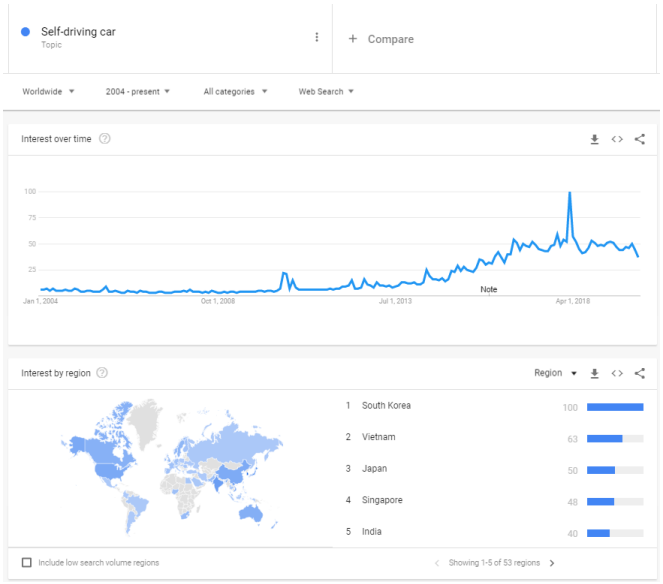
Figure 6: Worldwide interest in the topic "Machine Learning", 2004-19[58]



[57] Google Trends, https://trends.google.co.uk/trends/explore?q=%2Fm%2F0mkz
[58] Google Trends, https://trends.google.co.uk/trends/explore?date=all&q=%2Fm%2F0bs2j8q

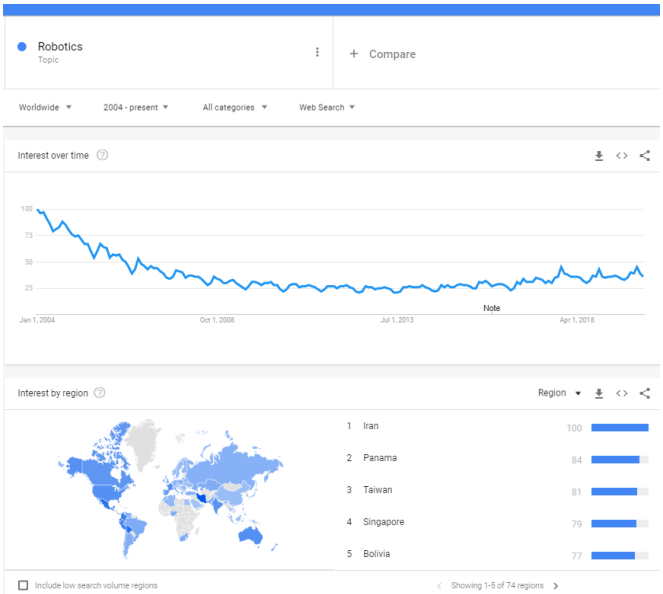
Interest in self-driving cars is also greater in Asia than in the West (Figure 7), though it is worth noting that the peak in March 2018 coincides with news coverage of a self-driving vehicle operated by Uber striking and killing a pedestrian in Arizona[59].

Figure 7: Worldwide interest in the topic "Self-driving Car", 2004-19[60]



Meanwhile, interest in robotics has declined since 2004. It is most prominent in searches from users in Asia and South America (Figure 8).

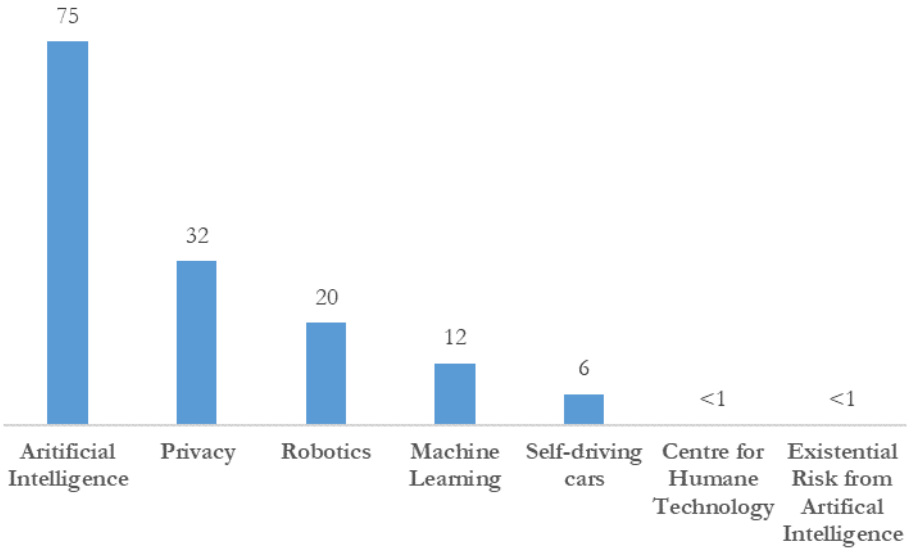
Figure 8: Worldwide interest in the topic "Robotics", 2004-19[61]



[59] Levin, S. and Wong, J.C. (2018), "Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian", *The Guardian*, https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe, accessed 7 January 2020
[60] Google Trends, https://trends.google.co.uk/trends/explore?date=all&q=%2Fm%2F0bs2j8q
[61] Google Trends, https://trends.google.co.uk/trends/explore?date=all&q=%2Fm%2F02p0t5f

Finally, a comparative analysis of worldwide Google search volume for these topics provides some clues as to their relative salience among members of the public over the last five years (Figure 9). One possible implication of this is that questions of design ethics and existential risk are an elite concern.

Figure 9: Relative salience of different technology topics among users of Google search, 2015-20 (0-100 index)



1.4 Responses from governments, institutions, and technology companies

Although the legal and regulatory environment is not the focus of this report, it warrants discussion in so far as it relates to underlying ethical issues. Governments and intergovernmental organizations (IGOs) have begun to explore and legislate on a broad range of these.

While much commentary on the European Union's General Data Protection Regulation (GDPR) has focused on its establishment of new rights for citizens such as the "right to be forgotten" and the duty it places on organizations to achieve "data protection by design", it also provides tools to ensure that organizations justify algorithmic decision-making[62]. Similarly, the Canadian government has issued a directive on automated decision-making that seeks to increase transparency, while the guidelines for AI R&D produced by Japan's Ministry of Internal Affairs and Communications aim to protect the interests of citizens[63]. Bias in algorithmic decision-making is also a concern of the UK government's Centre for Data Ethics and Innovation, along with targeted digital advertising, deepfakes, AI insurance, and so-called smart speakers[64]. In China, Article 8 of the Public Pledge of Self-Regulation and Professional Ethics for the Internet Industry commits tech companies to protecting the confidentiality of consumers' personal data, and to not using personal data for activities unrelated to the service being provided or in ways

[62] Gillis and Simons
[63] Deloitte Insights (2019), "Government Trends 2020", https://www2.deloitte.com/us/en/insights/industry/public-sector/government-trends.html, accessed 9 January 2020
[64] UK Government, Centre for Data Ethics and Innovation, https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation, accessed 9 January 2020

that are otherwise against the interests of users. Meanwhile, self-driving cars were the focus of guidelines issued by Germany's Ethics Commission on Autonomous Vehicles, which determined that in situations where human life is endangered, the car must act on utilitarian principles to minimize harm[65].

By contrast, technology companies' proactive efforts in digital ethics have been more narrowly focused. Many AI Ethics Boards have been established, but often the motivation for this has appeared to be executives' personal views on existential risk, rather than more immediate ethical considerations such as algorithmic fairness (see Figure 10). Concern about existential risk was the reason DeepMind required Google to create an AI ethics and safety board as a condition of its acquisition in 2014[66]. That the board closed one week after launching amidst criticism over the appointment of a board member with illiberal views on minority rights suggests a significant lacuna in its terms of reference, and hence in Google's thinking on AI ethics[67].

Figure 10: Selected tweets by Elon Musk on the existential risk from AI[68]

[65] Gershgorn, D. (2017) "Germany's self-driving car ethicists: All lives matter", https://qz.com/1061476/germanys-new-regulations-on-self-driving-cars-means-autonomous-vehicles-wont-compare-human-lives/, accessed 9 January 2020

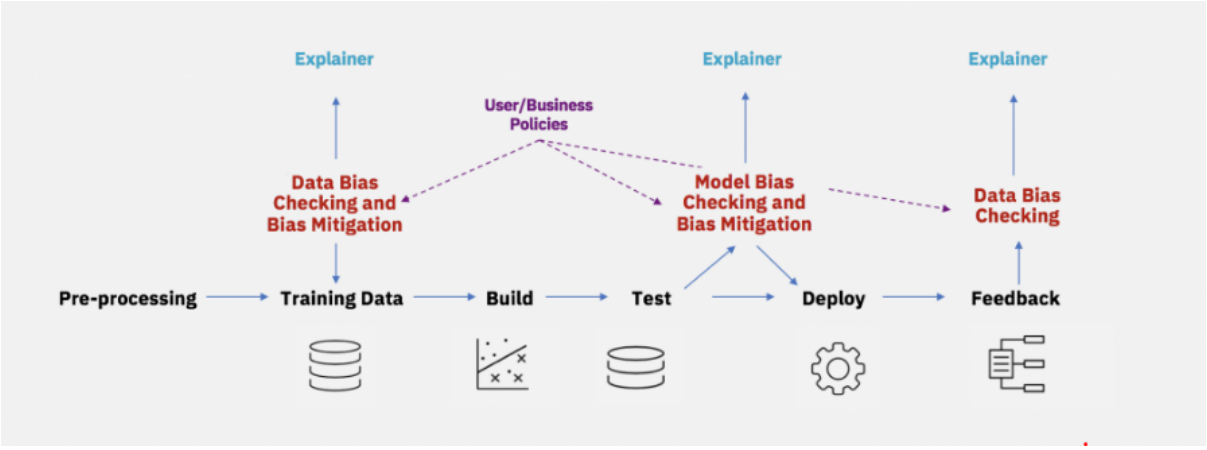
[66] Huffpost (2014), "Google's A.I. Ethics Board Might Save Humanity", https://www.huffpost.com/entry/googles-ai-ethics-board-might-save-humanity_n_5b573f66e4b01e373aac20b6, accessed 9 January 2020

[67] Piper, K. (2019), "Google cancels AI ethics board in response to outcry", *Vox*, https://www.vox.com/future-perfect/2019/4/4/18295933/google-cancels-ai-ethics-board, accessed 9 January 2020

[68] @elonmusk, Twitter, https://twitter.com/elonmusk/status/495759307346952192, https://twitter.com/elonmusk/status/871886151014940672, https://twitter.com/elonmusk/status/904639405440323585, accessed 9 January 2020

Technology companies' main engagement with mitigations to algorithmic unfairness has come in the form of resources for developers. These include Facebook's internal tool "Fairness Flow"[69], which measures how algorithms affect specific groups, and the "What-if tool" which is intended to help Google Cloud developers better identify biases in datasets and algorithms[70]. IBM has gone a step further with its open-source toolkit AI Fairness 360, which is designed to check for and mitigate unwanted biases embedded in model training data and algorithms. (Figure 11)[71].

Figure 11: Schematic of IBM's AI Fairness 360 toolkit



Meanwhile, the Partnership on AI, a non-profit think tank founded by Amazon, Facebook, Google, Microsoft, and IBM has addressed questions of job displacement in its research projects (in addition to questions of algorithmic transparency and fairness)[72]. Consistent with his hypothesis about the pace of advancement in AI, Elon Musk has supported calls for a Universal Basic Income (UBI), as have eBay founder Pierre Omidyar and Sam Altman, President of the Silicon Valley accelerator Y-Combinator, which numbers Airbnb, Stripe, and Dropbox among its alumni[73]. At a 2017 meeting of the World Economic Forum in Davos, Microsoft CEO Satya Nadella remarked that "We should do our very best to train people for the jobs of the future" while his counterpart at Salesforce, Marc Benioff, spoke of his fear that AI would engender "digital refugees"[74]. However, while job displacement may be a personal concern of technology

[69] Facebook (2018), "AI at F8 2018: Open frameworks and responsible development", https://engineering.fb.com/ml-applications/ai-at-f8-2018-open-frameworks-and-responsible-development/, accessed 9 January 2020
[70] Google (2018), "Introducing the What-If Tool for Cloud AI Platform models", https://cloud.google.com/blog/products/ai-machine-learning/introducing-the-what-if-tool-for-cloud-ai-platform-models, accessed 9 January 2020
[71] IBM (2018), "Introducing AI Fairness 360", https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/, accessed 9 January 2020
[72] The Partnership on AI, partnershiponai.org/about/, accessed 9 January 2020
[73] Clifford, C. (2017), "Y Combinator president and eBay founder join Elon Musk in addressing crisis of robots taking jobs", *CNBC*, https://www.cnbc.com/2017/02/13/y-combinator-president-and-ebay-founder-address-robots-taking-jobs.html, accessed 9 January 2020
[74] Kharpal, A. (2017) "Tech CEOs back call for basic income as AI job losses threaten industry backlash", *CNBC,* https://www.cnbc.com/2017/02/21/technology-ceos-back-basic-income-as-ai-job-losses-threaten-industry-backlash.html, accessed 9 January 2020

company CEOs and a focus for some of their philanthropic activity[75], there is less evidence of tangible action by technology companies to mitigate it.

IBM is also one of the circa 150 members of the International Data Spaces Association (IDSA), which brings together research organizations, industry, and technology companies including SAP, Huawei, and Deutsche Telekom. IDSA aims to establish the technical standards, architecture, and regulations needed for distributed data marketplaces. Although its focus is on secure data exchange between organizations, hypothetically its solution could form the basis for individuals to monetize their own data[76].

Turning to user interface design, Mark Zuckerberg appears to have absorbed criticisms levelled by design ethicists at social media companies. On an earnings call, he used the language of the "Time Well Spent" movement (the predecessor of the Center for Humane Technology) when explaining design changes to the Facebook and Instagram feeds[77]. These changes were intended to increase the quality and reduce the duration of user sessions, with an adverse impact on advertising revenues. Facebook has also announced a "pivot to privacy", suggesting that it will work towards making all private messages shared through its apps end-to-end encrypted by default[78]. At the same time, Google has followed Facebook's lead by incrementally improving the functionality which allows users to control how their personal data is shared across digital properties and used in targeted advertising[79].

[75] See for example Google.org, "Helping prepare for the future of work", https://www.google.org/helping-prepare-for-the-future-of-work/"
[76] International Data Spaces Assocation website, https://www.internationaldataspaces.org/the-principles/, accessed 24 February 2020
[77] Facebook (2017), "Facebook Q3 2017 Earnings Call". *Zuckerberg Transcripts*. 287. https://dc.uwm.edu/zuckerberg_files_transcripts/287
[78] Zuckerberg, M. (2019), "The Internet needs new rules. Let's start in these four areas". *Zuckerberg Transcripts*. 1008. https://dc.uwm.edu/zuckerberg_files_transcripts/1008
[79] See Google Safety Center Privacy Controls, https://safety.google/privacy/privacy-controls/

1.5 Discussion

The sudden proliferation of Ethics Boards at technology companies during 2018 and 2019 has led to accusations of "ethics washing" – that is, the disingenuous engagement with ethics as an external affairs strategy, in a manner analogous to the "greenwashing" marketing activities historically undertaken by companies with poor environmental records. As Meredith Whittaker points out, even Axon, which manufactures taser weapons, surveillance drones, and AI-enhanced police body cameras now has an Ethics Board. The establishment of such a board is clearly not by itself a sufficient response to the ethical issues raised by Axon's business activities. Unless Ethics Boards have the authority to veto product decisions and meaningfully hold executives to account for the ethical implications of their decisions, they risk being no more than "ethics theatre"[80]. This is also true of related moves by technology companies to publish ethical principles and appoint Chief Ethics Officers: without tangible evidence that they are driving ethical actions, they are likely to lack credibility, as the Salesforce case illustrates[81].

Case Study: Salesforce

Salesforce states that it has "a broader responsibility to society…to create technology that…**upholds the basic rights of every individual**" [emphasis added]

In 2018, it appointed Paula Goldman as **Chief Ethics and Humane Use Officer**. Goldman's remit is to:

- Understand the direct impact of Salesforce products on the world
- Create an ethical internal culture and product design process
- Advance the field through multi-stakeholder dialogues

Nevertheless, Salesforce faces ongoing criticism over its contract with the US Customs and Border Protection Agency, which implicates it in the **allegedly inhumane treatment of migrants** at the US's southern border.

It is not only technology companies which have been criticized for "ethics washing". A polemic by Thomas Metzinger, a member of the European Commission's AI expert group, described its Ethical Guidelines for Trustworthy AI as "lukewarm, short-sighted and deliberately vague. They ignore long-term risks, gloss over difficult problems ("explainability") with rhetoric, violate elementary principles of rationality and pretend to know things that nobody really knows"[82]. Of particular concern for Metzinger was the Commission's refusal to accept "red lines" on the use of autonomous weapons and on AI-based social credit scoring.

Overall responses to digital ethics issues might also be criticized on two other dimensions. Firstly, there appear to be areas where activity by governments and technology companies is somewhat misaligned with public concerns. Existential risk, for example, features disproportionately in technology companies' data ethics activities relative to the clear and present

[80] Crawford, K. and Whittaker, M. (2018) "How will AI change your life?" (Podcast), *Recode Decode*, https://www.vox.com/podcasts/2019/4/8/18299736/artificial-intelligence-ai-meredith-whittaker-kate-crawford-kara-swisher-decode-podcast-interview, accessed 10 January 2020
[81] Salesforce, "Ethical and Humane Use", https://www.salesforce.com/company/ethical-and-humane-use/; Cision (2018) "U.S. Customs and Border Protection Agency Selects Salesforce as Digital Modernization Platform", https://www.prnewswire.com/news-releases/us-customs-and-border-protection-agency-selects-salesforce-as-digital-modernization-platform-300608614.html; accessed 10 January 2020
[82] Metzinger, T. (2019) "Ethics washing made in Europe", *Der Tagesspiegel*, https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html, accessed 10 January 2020

danger of algorithmic injustice. Secondly, despite the global reach of the technology companies discussed in this report, there appears to be limited consideration of non-Western perspectives. There is less interest in data privacy in Asia and Latin America than in Europe, and more interest in technological job displacement – but the former has received greater attention in responses than the latter. If one accepts the proposition of Gry Hasselbalch and Pernille Tranberg that data ethics is the "new competitive advantage" for technology companies, actions in these underdeveloped areas would be a wise commercial move[83] – a question we return to in Part Three.

[83] Hasselbalch, G. and Tranberg, P. (2016), *Data Ethics: The New Competitive Advantage*, Publishare

2. The Moral Responsibilities of Technology Companies

2.1 The Necessity of Business Ethics

Do businesses have moral responsibilities at all? In the neoliberal paradigm, businesses' only duties are to act in the interests of their shareholders and to obey the law. Talk of "business ethics" is therefore nonsensical – a category mistake.

However, in practice this position is untenable for multinational corporations, as ethics and shareholder value are tightly intertwined. Journalists, NGOs, and activist shareholders routinely highlight issues such as poor treatment of workers and environmental degradation, while tax avoidance and regulatory arbitrage are increasingly regarded as unethical, irrespective of their legality. Historically, the involvement of corporations in catastrophes has materially damaged share prices, as BP's financial performance following the Deepwater Horizon oil spill underlines (Figure 12).

Figure 12: Impact of Deepwater Horizon oil spill (20th April 2010) on BP share price, 2009-13[84]

While short-term falls in share price following catastrophes are explained by the magnitude of financial loss and the number of fatalities, management responsibility for safety lapses has been shown to be the key driver of longer-term reductions in shareholder value[85]. Three examples of this are shown in Figure 13: the performance of Johnson and Johnson shares after poisoning deaths resulting from tampering with capsules of the painkiller Tylenol in 1982; of Exxon shares following the 1989 spill from the Valdez oil tanker; and of shares in the chemical company Union Carbide after the toxic gas leak at Bhopal in 1984, which caused up to 16,000 deaths.

[84] Charting: Hargreaves Lansdown, https://www.hl.co.uk/shares/shares-search-results/b/bp-plc-ordinary-us$0.25/share-charts
[85] Knight, R.F. & Pretty, D.J. (1997) "The Impact of Catastrophes on Shareholder Value", *Oxford Executive Research Briefings*, Templeton College, Oxford, Available at: http://eternity.websurgeon.ca/papers/whitepapers/sedgwickreport.pdf

Figure 13: Selected examples of catastrophes with long-term adverse impact on shareholder value[86]

Catastrophe	Cumulative Abnormal Returns	
	At 50 Trading Days	At 6 Calendar Months
Johnson and Johnson – Tylenol	(10%)	(18%)
Exxon – Valdez	(15%)	(18%)
Union Carbide – Bhopal	(29%)	(29%)

2.2 Ethical Traditions and Universal Norms

Ethics is therefore a significant question for leaders of multinational corporations, regardless of their personal perspectives on the subject. But how are they to approach ethics in practice, given the different ethical traditions around the world? Some of the dominant ethical traditions which multinationals may expect to encounter are as follows:

- Islamic Tradition
- Confucian Tradition
- Western / Judaeo-Christian Tradition
- Indian Tradition
- Buddhist Tradition

Even in this far-from-exhaustive list, there are significant points of divergence on matters of business ethics. Trust-based relationships may be viewed as more or less important than legal contracts. Favouritism towards family members may be regarded as a virtue or as a vice. Furthermore, divergence also occurs *within* ethical traditions. For example, within the Western tradition, America emphasizes the virtue of self-reliance, while Europe emphasizes the duty to look after the needy. In the Islamic tradition, law is interpreted based on local context, leading to variations in ethical conduct[87].

Examples of "The Golden Rule" in Confucian and Western ethical traditions

'Tsze-kung asked, saying, "Is there one word which may serve as a rule of practice for all one's life?" The Master said, "Is not RECIPROCITY such a word? What you do not want done to yourself, do not do to others.'"

– Confucius, The Analects, Book 15

'All things whatsoever ye would that men should do to you, do ye even so to them'

– Matthew 7:12

However, there are also significant points of convergence. Fulfilling duties and treating people as you would wish to be treated are universally regarded as virtues. Harming people, stealing, telling lies, and committing fraud are universally regarded as vices. It has therefore been possible to determine universal minimum ethical standards for multinational business activities, based on these points of convergence. These standards are summarized in Figure 14.

[86] Knight & Pretty (1997)
[87] Hendry, J. (2013), "Ethical Cultures and Traditions", Second Edition, available at http://johnhendry.co.uk/wp/wp-content/uploads/2013/05/Ethical-cultures-and-traditions.pdf

Figure 14: Minimum ethical standards in multinational business[88]

What	How
Respect for human dignity and basic rights	• Treating employees, customers, and suppliers as people with intrinsic value (not as means to an end) • Producing safe products and services • Ensuring workplaces are safe • Upholding individuals' rights to education and an adequate standard of living
Good corporate citizenship	• Supporting essential social institutions (e.g. economic system, education system) • Working with governments to protect the environment

While, as we have seen, digital technology raises new questions which cannot always be addressed by applying standards developed for industries such as pharmaceuticals, oil, and chemicals, we are optimistic that a comparable "overlapping consensus" on data ethics can be achieved[89]. The following case study brings this to life.

2.3 Case Study: Shanghai Roadway

Shanghai Roadway ("Roadway") was a China subsidiary of the global information services firm Dun & Bradstreet ("D&B"). Based in New Jersey and with 5,000 employees, D&B provides data and analytics to approximately 90% of Fortune 500 companies. Prior to a private equity buy-out in 2019, the company was listed on the New York Stock Exchange, and reported revenues of $1.7bn in 2017[90]. D&B acquired Roadway in 2009 as part of a strategy to grow its business in the Chinese market. Like many other divisions of D&B, Roadway marketed data about businesses and consumers to lenders and other commercial clients, generating revenues of $23m in 2011[91].

It was the data collection practices used by Roadway which were to result in controversy. In common with the business model of other information services firms discussed in Part One of this report, Roadway sourced data on individuals to populate its databases through commercial relationships with banks, insurance companies, real estate agents, and telemarketing companies. The company held personal information (including income levels, jobs, and addresses) on

[88] Adapted from Donaldson, T. (1996), "Values in Tension: Ethics Away from Home", *Harvard Business Review*, https://hbr.org/1996/09/values-in-tension-ethics-away-from-home
[89] The term "overlapping consensus" was coined by the philosopher John Rawls in *A Theory of Justice* (1971). It means that people can agree with the same laws for different reasons.
[90] Dun & Bradstreet company website, https://www.dnb.com/about-us.html
[91] Chu, K. (2013) "Dun & Bradstreet Fined, Four Sentenced in China", *Wall Street Journal*, https://www.wsj.com/articles/SB10001424127887323482504578230781008932240, accessed 14 January 2020

approximately 150 million Chinese citizens. Records from this database were sold for RMB1.5 (~$0.23) each to companies as marketing & sales leads.

In September 2012, the Shanghai District Prosecutor charged Roadway and five former employees with illegally obtaining private information belonging to Chinese citizens. The court fined Roadway RMB1m ($160,000) and sentenced four of the former employees to two years in prison. The basis for the court's decision was a 2009 amendment to the Criminal Law, which made it illegal for companies in the financial services, telecommunications, transport, education, and healthcare industries to obtain or sell a citizen's personal information[92]. Following the judgment, Roadway ceased trading and D&B reported itself to US regulators. In April 2018, D&B agreed to pay a $9m fine to resolve charges under the Foreign Corrupt Practices Act. These charges related to payments made to third-party agents who had procured data on Roadway's behalf, and to bribery of government officials by another China subsidiary of D&B, HDBC, to facilitate access to government-held personal data[93].

The significance of the Roadway case is that it demonstrates an "overlapping consensus" on a tangible question of data ethics. Chinese and US prosecutors agreed that Roadway had acted unethically, and sanctioned D&B accordingly. However, they did not agree on what aspect of Roadway's conduct was unethical: in China, it was its business model of collecting and selling personal data on Chinese citizens without informed consent; in the US, it was the practice of making off-book payments to intermediaries and government officials. The former is ethically acceptable in the US – indeed, there is an entire $2.6bn industry, digital lead generation, devoted to it[94]. In the context of Chinese norms of gift-giving, the latter can be perceived as an honourable way of investing in a long-term business relationship, rather than an unethical instance of bribery.

2.4 Philosophical Ethics

There are three main theories of philosophical ethics: deontology, utilitarianism, and virtue ethics, associated with the thought of Immanuel Kant, Jeremy Bentham, and Aristotle respectively. We encountered deontology and utilitarianism in the context of the Trolley Problem, discussed in Part One of this report. Kant would *not* pull the lever to divert the runaway train, on the basis that it is one's duty not to kill, regardless of the extenuating circumstances. Bentham, on the other hand, would calculate what action would lead to the

[92] Inside Privacy (2013), "Dun & Bradstreet Reportedly Fined RMB $1 Million for Illegally Obtaining Personal Information in China; Four Employees Imprisoned", https://www.insideprivacy.com/international/dun-bradstreet-reportedly-fined-rmb-1-million-for-illegally-obtaining-personal-information-in-china/, accessed 14 January 2020
[93] Volkov, M. (2018), "Dun and Bradstreet Pays $9 Million for FCPA Violations in China", https://blog.volkovlaw.com/2018/05/dun-and-bradstreet-pays-9-million-for-fcpa-violations-in-china/, accessed 14 January 2020
[94] Statista, "Digital lead generation ad spend in the US, 2019-2023", accessed 14 January 2020

greatest overall happiness, and decide to pull the lever, on the basis that the happiness created by saving five lives would more than offset the happiness destroyed by causing one death.

With the Roadway case in mind, it is hard to see how either a deontological or utilitarian approach to data ethics is feasible for multinational technology companies. Deontology requires determinations about whether actions are right or wrong in and of themselves, leaving no room for reconciling variations in norms in different regions of the world. Utilitarianism, meanwhile, requires weighing up the consequences of different actions. Even in the relatively simple example of Roadway's procurement of personal data, it is very difficult to make this calculation. What proportion of the Chinese citizens whose data was traded by Roadway experienced harm, whether through unwanted phone calls or messages, or in more serious forms? What economic benefits accrued to shareholders and employees of Roadway's clients from using the data, and how much of those benefits flowed on to their own clients, shareholders, and employees? How are these harms and benefits to be quantified? More complex questions of data ethics – in relation to the development of recidivism algorithms, for example – are even more resistant to utilitarian calculus.

This brings us to virtue ethics. For Aristotle, ethics should be less concerned with specific actions than with the character of the person (or organization) who carries them out. In the Trolley Problem, whether or not it is ethical to pull the lever therefore depends on the particular circumstances of the person faced with the choice: pulling the lever might express the virtue of bravery, or the vice of megalomania. When applied to data ethics, virtue ethics suggests the best question technology companies can ask themselves is: "What kind of company should we be?" In the Roadway case, D&B might have asked, "Do we want to be the sort of company that covertly gains access to personal data by making off-book payments?" Posing this question helps identify what may be a universal virtue of technology companies: transparency. Acting transparently would have required both making it clear to consumers how data about them might end up in a commercially-available marketing database, and eschewing extra-contractual arrangements with suppliers and intermediaries.

2.5 Case Study: Cloudflare and 8Chan

Cloudflare's decision to ban 8Chan as a client provides an example of virtue ethics in practice. Cloudflare is a San Francisco-based web infrastructure company which provides cloud security, protecting websites from Distributed Denial of Service and other cyberattacks. It supports over 19 million websites for clients including IBM, Thomson Reuters, and Zendesk[95].

Until 2019, another of Cloudflare's clients was 8Chan, a message board associated with white-supremacist and neo-Nazi ideologies. Historically, Cloudflare regarded itself as a neutral utility service, and therefore did not make judgements about its clients based on the content of

[95] Cloudflare company website, accessed 15 January 2020

their websites. Such a position could be justified from a deontological or a utilitarian perspective: one could argue that it is simply wrong for a private company to censor speech, or that removing a client like 8Chan might lead (say) to pressure from governments to remove websites belonging to unfavoured minorities[96]. However, when it became clear that the perpetrator of the mass shootings in El Paso, Texas in August 2019 had posted a manifesto on 8Chan, Cloudflare's CEO Matthew Prince changed his stance and terminated 8Chan as a client. In a blogpost on Cloudflare's website Prince wrote,

> We do not take this decision lightly. Cloudflare is a network provider. In pursuit of our goal of helping build a better internet, we've considered it important to provide our security services broadly to make sure as many users as possible are secure, and thereby making cyberattacks less attractive — regardless of the content of those websites. Many of our customers run platforms of their own on top of our network. If our policies are more conservative than theirs it effectively undercuts their ability to run their services and set their own policies. We reluctantly tolerate content that we find reprehensible, but we draw the line at platforms that have demonstrated they directly inspire tragic events and are lawless by design.[97]

Cloudflare, it seems, wants to be the kind of company that upholds net neutrality as far as is possible without undermining the rule of law. It is also worth noting that by writing publicly about his decision-making in all its nuance and complexity, and by giving interviews to journalists about it, Prince exhibited the virtue of transparency.

2.6 Data ethics in new product development

As another example of how a technology company might apply virtue ethics in practice, let us consider a hypothetical provider of telecoms infrastructure and hardware. Seeking growth, such a provider might consider diversifying into Value-Added Services (VAS) such as messaging applications, location services, IoT analytics, and mobile advertising. All such services necessarily involve the collection and storage of personal data: even end-to-end encrypted messaging produces metadata which could contribute to the identification of individual users; even business-to-business IoT applications could capture data on individual employees or customers via sensors. At the same time, storing personal data flowing from VAS creates the risk that data may be misappropriated by malevolent actors and used to harm individuals in the ways described in Part One of this report.

In short, providing VAS creates new risks that individuals will be exposed to harm. In addition, as data protection legislation typically lags advances in digital technology, the law cannot be consistently relied on for guidance. One potential response to these risks might be to adopt best-practice data protection standards such as:

- **Transparency**: making it clear to individuals what personal data is being collected by VAS and what it will be used for

[96] Roose, K. (2019a), "Why Banning 8chan Was So Hard for Cloudflare: 'No One Should Have That Power'", *New York Times*, https://www.nytimes.com/2019/08/05/technology/8chan-cloudflare-el-paso.html, accessed 15 January 2020
[97] Prince, M. (2019), "Terminating Service for 8Chan", https://blog.cloudflare.com/terminating-service-for-8chan/, accessed 15 January 2020

- **Consent**: ensuring individuals give informed consent for data collection and are able to withdraw (or condition) their consent in future
- **Data minimization**: collecting only the data which is required for the functionality of the VAS
- **De-identification**: responsibly managing the risk of re-identification where behavioural data and IoT data are collected
- **Security**: responsibly managing the risk of data breaches

Further issues arise if the data produced by providing VAS is put to secondary uses, such as using location data to target advertising, or packaging IoT data from smart home devices as an analytics product. Since secondary uses are often unknown at the point of data collection, they are in tension with the principle of consent. Similarly, big data analytics is in tension with the principle of data minimization, as it is based on the serendipitous discovery of correlations in very large datasets.

At the same time, development of VAS may lead to unintended consequences. Location services have been used by perpetrators of domestic abuse to track down their victims[98]. Livestreaming services have been instrumentalized to broadcast footage of self-harm, suicide, and mass killings[99]. Encrypted messaging services have been used to circulate images of child sexual abuse and plan acts of terrorism[100].

Do the unintended consequences imply that it would be unethical for the telco to develop new products in these areas at all? Do the best practices imply that VAS can only be ethical if their scope is tightly circumscribed? Surely not: to argue this would be to wilfully ignore the beneficial consequences for large numbers of users and clients of providing the services, and to foreclose on opportunities for data analytics to create public value (imagine, for example, that smart home analytics provided new insights enabling major advances in domestic energy conservation). Instead, asking "What kind of company should we be?" can help strike a balance between pursuing innovation, growth, or profitability at any cost and the inertia that follows from disproportionate risk-mitigation.

[98] Valentino-DeVries, J. (2018), "Hundreds of Apps Can Empower Stalkers to Track Their Victims", *New York Times*, https://www.nytimes.com/2018/05/19/technology/phone-apps-stalking.html, accessed 15 January 2020
[99] Tanakasempipat, P. and Thepgumpanat, P. (2017). "Thai man broadcasts baby daughter's murder live on Facebook". *Reuters,* https://www.reuters.com/article/us-thailand-facebook-murder-idUSKBN17R1DG; Roose, K. (2019b), "A Mass Murder of, and for, the Internet", *New York Times,* https://www.nytimes.com/2019/03/15/technology/facebook-youtube-christchurch-shooting.html; accessed 15 January 2020
[100] Newton, C. (2019), "The big disturbing problem that could help end encryption", *The Interface,* https://www.getrevue.co/profile/caseynewton/issues/the-big-disturbing-problem-that-could-help-end-encryption-201636, accessed 15 January 2020

3. Data Ethics Recommendations

3.1 Existing data ethics best practice

We begin this part of the report with a brief summary of existing data ethics best practice. This is not presented as a checklist for technology companies: that would be inconsistent with the virtue ethics-based approach we have advocated. Instead it should be seen as stimulus for ethical reflection on "the kind of company we should be".

- Expand the scope of the compliance function to "Ethics and Compliance", acknowledging that all companies have broader social responsibilities than complying with the law and mitigating regulatory risk, and that the kind of power technology companies have heightens these responsibilities.
- Establish a group with oversight responsibility for data ethics, with a formal role in the company's governance, including veto rights over product decisions. Without such rights, ethics boards struggle to transcend "ethics theatre".
- The scope of governance initiatives should not be arbitrarily limited to AI (and certainly not to existential risk). As this report makes clear, data ethics questions run broader and deeper, and many are more immediate than the threat of machine superintelligence.
- Support collaborative data ethics initiatives (for example: the Partnership on AI and the International Data Spaces Association), and contribute technology expertise to consultations by governments and IGOs on questions of data ethics.
- Provide transparency to individuals over how their personal data is collected, stored, and used, and offer tools enabling them to exercise control over it (for example: Google's Privacy Center).
- Review the suitability of "engagement" metrics as business targets, as they may be misaligned – or even expressly at odds – with users' wellbeing.
- Where practical questions of data ethics are complicated by differences in ethical traditions and norms (as in the Shanghai Roadway case), "overlapping consensus" may be achievable and should be pursued.
- Use an ethical framework when evaluating whether to honour or terminate client relationships which may be enabling unethical practices (for example: Salesforce and the US Customs and Border Protection Agency)
- Conduct deliberation of data ethics questions in public as far as possible (as in the case of Cloudflare and 8Chan)
- Approach new product developments with due consideration for both data-related risks and potential unintended consequences – sometimes referred to as adopting "the precautionary principle", as opposed to an approach of "permissionless innovation"[101].

[101] For a discussion, see Rosner, G. and Thierer, A. (2018), "The Precautionary Principle vs Permissionless Innovation", *Governing the Internet of Things*, American University Internet Governance Lab, https://www.iotprivacyforum.org/2018/03/14/governing-the-internet-of-things/, accessed 15 January 2020

In our view, this summary reflects a limitation of current discourse on data ethics: it emphasizes risk mitigation over positive actions. We therefore conclude the report with some fresh ideas for ways multinational technology companies can act to maximize their ethical impact.

3.2 Co-creating the jobs of the future

Automation is not the only driver of disruption to work: as the emergence of the term "the gig economy" indicates, the labour market is becoming more flexible and more volatile. Although there has been no shortage of corporate research initiatives on the "future of work", most tangible interventions have come from startups and social enterprises responding to the increasing insecurity of employment already being experienced by many workers[102]. There is therefore an opportunity for the kind of technology companies who seek to enable human flourishing to proactively fashion the jobs of the future – that is, jobs in which humans and machines are not in competition, but instead collaborate in a way that both enhances performance and enriches human life.

Precedents for this form of working already exist. Machine-learning algorithms can be as effective as experienced doctors at diagnosing some forms of cancer[103], pointing to a future for medical professionals in which human clinical experience is complemented by AI's strengths in pattern-recognition and data processing speed[104]. However, opportunities also exist in less specialized types of work, where far more people are employed. One example is food ordering and delivery, where technology companies such as Deliveroo and Alibaba's Ele.me have created high-growth businesses by combining mobile apps, logistics, and casualized labour. Ethnographic research has revealed a somewhat adversarial relationship between human couriers and the algorithms used to assign jobs and plan routes. In both face-to-face conversations and WhatsApp groups, couriers collaborate in an attempt to reverse-engineer the algorithm, iteratively testing strategies intended to optimise their earnings. Meanwhile, isolated in a head office environment where they never interact with the couriers, the algorithm's developers build new features based on assumptions about user needs which the couriers know to be flawed[105].

A "co-creation" approach offers the potential to re-design the relationship between human workers and AI to be more harmonious and more economically productive. The process of co-creation involves bringing together representative groups of stakeholders to work collaboratively on a solution, with independent facilitation. In the case of food delivery, these stakeholders would be the couriers, the algorithm developers, and users of the apps. Although co-creation is usually associated with the development of new consumer products and services, it has been proposed as an approach to mapping the future of work at the Mondragón Corporation in the Basque region, where automation and the need for a green transition are creating uncertainty for

[102] Leadbeater, C. (2019), "The RSA Future Work Awards — meeting anxiety with innovation", *The RSA*, https://medium.com/@thersa/the-rsa-future-work-awards-meeting-anxiety-with-innovation-84a1ed1d6b0a
[103] See for example McKinney, S.M., Sieniek, M., Godbole, V. et al. (2020) "International evaluation of an AI system for breast cancer screening". *Nature* 577, 89–94, doi:10.1038/s41586-019-1799-6
[104] For a discussion, see Fry (2018)
[105] Perrig, L. (2019) "Matching users' and developers' beliefs: The algorithmic management of uncertainty". Presentation at *Connected Life: Data & Disorder*. London School of Economics

large numbers of manufacturing and construction workers[106]. Technology companies could undertake co-creation processes like this with their own employees in fields which are at risk of technological displacement, with the employees of their subcontractors and suppliers, or with people whose employability is threatened by their products. Millions of workers in fields like customer services, social media content moderation, and device manufacture and assembly could stand to benefit from new kinds of jobs designed through co-creation initiatives.

3.3 Opening Data

Technology companies with data-driven business models tend to reject the assertion that the services they provide to users are worth less than the user data they monetise, and have therefore not explored the idea of compensating users for data. In our view, there are philosophical barriers to treating data as private property[107], as well as practical barriers to paying users for data, as Lanier has proposed. Nevertheless, the kind of technology companies who want to promote public good by enabling scientific progress and social justice could consider sharing data value in a different form. Specifically, we propose that technology companies make a greater proportion of the data they have collected openly available for researchers, policymakers, and ordinary members of the public to re-use. The architecture and standards proposed by the IDSA (discussed in Part One) could provide the technical basis for opening up more data in a secure and trusted manner.

What sort of data do we mean and what use-cases do we have in mind? *Internet search data* provides a helpful example. Google already makes limited search data available through Google Trends – it was this tool which enabled us to measure the salience of different data ethics issues to the public over time and in different world regions in Part One. This data is demonstrably valuable for academic research, notably in public health, where it has been used to analyse the symptoms of fibromyalgia[108] and the seasonality in domestic violence[109], and to forecast transnational migration flows[110] and the spread of infectious diseases[111]. However, Google Trends provides very little data on search term variations – that is, what exactly users search for when they are searching on a particular topic. It is deep search term variation data which holds untapped insights into public opinion, attitudes, preferences, needs, desires, and behaviours. As

---

106 Agirre Lehendakaria Center for Social and Political Studies (2019), "Mondragon will count on Mariana Mazzucato in its commitment to social innovation", https://agirrecenter.eus/en/news/mondragon-will-count-on-mariana-mazzucato-in-its-commitment-to-social-innovation, accessed 17 January 2020

107 See Prainsack, B. (2019), "Logged out: Ownership, exclusion and public value in the digital data and information commons". *Big Data & Society*, 6(1): 2053951719829773.

108 Bragazzi, N.L., Amital, H., Adawi, M. et al. Clin Rheumatol (2017) "What do people search online concerning the "elusive" fibromyalgia? Insights from a qualitative and quantitative analysis of Google Trends" *Clinical Rheumatology* 36: 1873. https://doi.org/10.1007/s10067-017-3665-y

109 Koutaniemi, E. M., & Einiö, E. (2019). Seasonal variation in seeking help for domestic violence based on Google search data and Finnish police calls in 2017. *Scandinavian Journal of Public Health*. https://doi.org/10.1177/1403494819834098

110 Sanliturk, A.E. (2019), "Search for a New Home: Use of Google Trends Data to Capture Migrant Mobility", in *Digital Demography in the Era of Big Data*, Instituto de Estadística y Cartografía de Andalucía, 6-7 June 2019

111 Watad, A., Watad, S., Mahroum, N., Sharif, K., Amital, H., Bragazzi, N. L., & Adawi, M. (2019). "Forecasting the West Nile Virus in the United States: An Extensive Novel Data Streams-Based Time Series Analysis and Structural Equation Modeling of Related Digital Searching Behavior". *JMIR public health and surveillance*, *5*(1), e9176. doi:10.2196/publichealth.9176

much of this data has negligible value to advertisers – think of the hundreds of millions of searches for state benefits, for example – there is little commercial reason for technology companies to hoard it. It is not only search engines like Google, Baidu and Bing which are able to capture internet searches: web browsers, browser extensions, anti-virus software applications, and internet service providers are too. Making more search data openly available – in de-identified and aggregated form – is one way that technology companies can promote public good. The same argument can be made about de-identified and aggregated *location data*, *IoT data*, and *clickstream data* (behavioural data about which websites users visit and in what sequence).

Regrettably, however, this is not the prevailing trajectory. Google recently deprecated Google Correlate, which enabled researchers to see what search terms were correlated with their own time-series datasets. Similar trends are apparent in relation to *social media data* from Facebook, Twitter, and Youtube. Motivated in part by calls for stronger controls to assure users' privacy in the wake of the Cambridge Analytica scandal, researchers' access to datasets is increasingly constrained – a development sometimes referred to as "The APIcalypse". Beyond academia, open-source intelligence organizations such as Bellingcat, which rely on access to social media data for their investigations into potential war crimes and humans rights violations, have found their work hampered by the deletion of video archives by Youtube and the withdrawal of tools like Facebook Graph Search and Google Earth Panoramio[112].

It is likely that technology companies are simply ignorant of many of the beneficial uses to which the data they already share is put. By definition, they cannot imagine the new beneficial uses which might be found for data they are yet to return to the public domain. These could easily include scientific advances, medical breakthroughs, criminal prosecutions, and public policy innovations. Does an ethical technology company seek to unlock those opportunities, or to foreclose them?

[112] Dubberley, S. (2019), "How Facebook's sudden change hinders human rights investigations", https://www.amnesty.org/en/latest/news/2019/06/how-facebooks-sudden-change-hinders-human-rights-investigations/, *Amnesty International*, accessed 17 January 2020