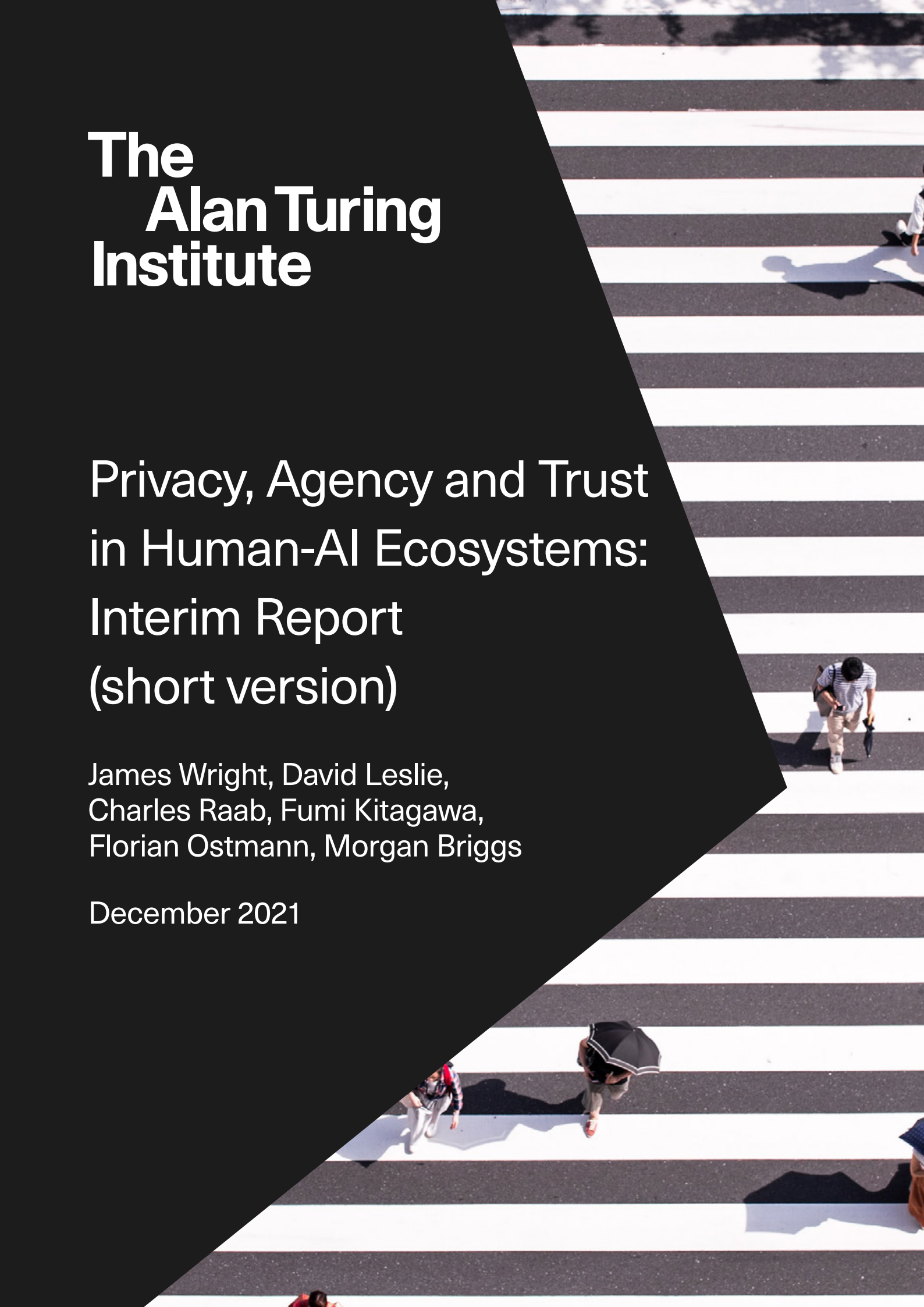
The Alan Turing Institute

Privacy, Agency and Trust in Human-AI Ecosystems: Interim Report (short version)

James Wright, David Leslie,
Charles Raab, Fumi Kitagawa,
Florian Ostmann, Morgan Briggs

December 2021



Executive Summary

PATH-AI (Privacy, Agency and Trust in Human-AI Ecosystems) is a collaborative and multidisciplinary research project involving The Alan Turing Institute, the University of Edinburgh, and the RIKEN research institute in Japan. The project is funded by UK Research and Innovation and the Japan Science and Technology Agency.

The aim of PATH-AI is to examine how the three interrelated values of privacy, agency, and trust are entangled in very different cultural contexts in relation to AI and other data-driven technologies, and how a focus on these values can inform the ongoing shaping of the international landscape for AI ethics, governance, and regulation.

This short-form version of the interim report presents an overview of the PATH-AI project, a brief discussion of the three key concepts of privacy, agency, and trust, as well as interim results from interviews and surveys conducted in the UK and Japan with 95 stakeholders comprising members of the general public as well as experts working on the implementation and/or ethics and governance of AI and other data-intensive technologies in healthcare. The full report is due to be published in early 2022.

Some of our key findings include the following:

- The concepts of privacy, agency, and trust are understood rather differently in the UK and Japan. Very broadly, in the UK, the liberal tradition of the autonomous and self-possessing individual has tended to predominate, with greater emphasis on individual rights and freedom to pursue one's personal self-interest. In Japan, these values have tended to be understood in a more relational, less absolute way, taking greater account of interdependent social relations and contextual identities.
- However, cultures are neither bounded nor monolithic, and it is important not to suggest simplistic or essentialised binary oppositions of East vs West. There are also many correlations and convergences between the UK and Japan driven by global flows of capital, information, labour, technologies, and neoliberal ideology, as well as by the imperative for interoperable legal and regulatory frameworks and standards to facilitate international trade.
- Across the interviews and surveys conducted with both UK and Japanese research participants, a thread that connected privacy, agency, and trust in the context of emerging AI and data-intensive digital technologies was the sense of growing asymmetries – of data, informed choice, resources, capabilities, and ultimately power – between users, governments, and companies.
- These asymmetries were fuelled by black-box digital tools, apps, and algorithmic systems, often developed and deployed in a top-down, paternalistic manner by tech companies and governments alike. This left many respondents feeling confused about what data was being collected about them and how it was being used, feeding feelings of disempowerment and distrust, with little participatory parity or agency.

- While the recent trend to greater anxiety about privacy seems to reflect a profound and growing societal distrust of both governments and technology companies, many experts argued that the implementation of data protection legislation has so far tended only to confuse citizens while not preventing companies from collecting ever-greater amounts of personal data.
- Respondents in both the UK and Japan were concerned that citizens, government, and regulatory and legal systems were simply not able to keep up with the rapidly increasing complexity of technological developments.
- Differences identified between the views of respondents in the UK and Japan included contrasting concerns reflecting differing healthcare systems. In the UK, many interviewees worried about tech companies leveraging their greater capacity for data collection and analysis to privatise the NHS by the back door. In Japan, the larger concern was that the country was falling behind less scrupulous global competitors due to siloed and inefficient bureaucratic structures and an overly cautious attitude from the private sector.
- There was also a notable difference in how AI and related technologies were viewed, with Japanese respondents more optimistic and seemingly more comfortable with the idea of, for example, interacting with robots.
- While UK participants worried that emerging technologies might cause harm due to inadequate design and implementation, many Japanese participants seemed more worried that they would work *too well* or become too powerful, introducing dangers of lack of control or the creation of a future society ruled by automated decisions and action that would be less easily contested.
- Across both Japan and the UK, experts and members of the public called for greater public education and far clearer communication about these increasingly complex technologies. They also called for more public consultations and meaningful participation in governance and regulation. Several experts argued for the need to embrace a collective and more horizontal approach that struck a new reconciliation between individual rights and public benefits.

The next stage of the PATH-AI project will build on this foundation by developing and piloting a methodology for the intercultural co-design of a framework for more ethical and equitable human-AI ecosystems.

Funding

This work was supported by the Economic and Social Research Council [grant number ES/T007354/1, Principal Investigator: Dr David Leslie] and the Japan Science and Technology Agency.

Acknowledgements

We would like to thank all the research participants in the UK and Japan who very generously gave their time to share their views. We would also like to express our appreciation to Jane Butler and Dame Philippa Russell from Carers UK, and Julien Danero Iglesias from Camden Council, who kindly helped publicise our call for interviewees. Finally, a special thanks to the members of the PATH-AI team in Japan – Dr Nakagawa Hiroshi, and Professors Ohya Takehiro, Narihara Satoshi, and Sakura Osamu – for their insights, inputs, and research that have helped inform this report.

Contents

Introduction

As the use of data-driven technologies, including artificial intelligence (AI), continues to expand in scope and scale, the demand for governments to find effective ways to regulate them is growing in kind. AI is heralded for enabling a new and revolutionary generation of powerful tools with vast potential to improve the economy and productivity, public services, health and wellbeing, and the environment. Yet the need for strong governance regimes to rein in the adverse individual and societal impacts of AI arises from the equally wide spectrum of harms that these technologies may engender.

At the individual level, the irresponsible use of AI systems can harm the dignity, autonomy, physical and mental integrity, and material wellbeing of the people who are impacted by them. Unbounded automation, intrusive personal profiling, and opaque algorithmic decision-making can leave individuals feeling disempowered, dehumanised, objectified, exposed, and manipulated. Widening forms of behavioural nudging by algorithms can have transformative effects on a person's inner life, undermining their sense of agency, and curtailing possible paths to their full self-formation and flourishing.

At the interpersonal level, crucial human connection, trust, and empathy may be lost through ever broader forms of algorithmic control. The increasing presence of biometric surveillance in the workplace and in public spaces can discourage or prevent citizens from exercising their freedoms of assembly and association, removing the protection of anonymity and having a chilling effect on social cohesion and democratic participation. Moreover, as other forms of automated social management continue to multiply, they threaten to weaken the connective tissue of solidarity. This is beginning to surface in the widespread deployment of algorithmic labour and productivity management tools, as well as in the rise of automated welfare systems and poverty management regimes, where predictive computational models are being used to allocate social services, prevent and prosecute criminal behaviour, and to determine individual risks of need and harm. These kinds of algorithmic decision system, as Virginia Eubanks writes, have the potential to "hide poverty from the professional middle-class public and give the nation the ethical distance it needs to make inhuman choices" (Eubanks 2018, 13). Such technological affordances run the risk of reframing the shared political responsibility of deciding what a fair and equitable society should look like as system-engineering problems to be solved by predictive analytics and by instrumentalised administrative techniques.

At the societal level too, the proliferation of AI technologies is posing significant risks of harm. The widespread use of data-driven predictive models in parts of society where historical legacies and patterns of discrimination, inequity, and bias are prevalent has so far shown signs of deepening such injustices rather than helping overcome them. Drawing insights from existing data, machine learning models, when they work reliably, make predictions about people's behaviour that by definition replicate the social and cultural patterns of the past—regardless of whether these patterns are inequitable or discriminatory.

Moreover, in current digital information and communication environments, the predominant steering force of social media and search platforms makes use of opaque computational methods of relevance-ranking, popularity-sorting and trend-predicting to manufacture digital publics largely devoid of active participatory social or political choice. Rather than being guided by the political will of citizens achieved through public discussion and deliberation, this vast mesh of connected digital services shapes these publics in accordance with the drive to target, capture, harvest and monetise individual attention (see, for example, Hao 2021). As this manufacturing of digital publics is pressed ever more into the service of profit-seeking, democratic agency may be increasingly supplanted by behavioural manipulation at scale. Combined with complementary dynamics of wealth polarisation and rising inequality, such a reduction in social capital and solidarity is already feeding the crisis of social and political polarisation, the widespread kindling of societal distrust, and the animus towards consensus-based science that have come to typify contemporary post-truth contexts.

Governments so far have not appeared equal to the task of remediating – let alone anticipating and preventing – this variety of potential harms. The transnational and global nature of the problems, and of the continuous flow of data and AI-enabled applications, has seemed to put their redress beyond the reach of individual nation-states and democratic decision-making. Questions of how and why data should be generated about us, whether by governments or corporations, in what ways this data should be shared and used, and how much we should be informed about these processes, are becoming increasingly pressing. Yet our conventional mechanisms of legal and regulatory recourse have so far appeared more atrophied than muscular. Beyond the question of states' capability to remediate, there is also the question of political will, as governing elites and those who supply the technologies have tended to promote, instead, the advantages of using AI in policy and administration and to pay less attention to the negative side.

Nevertheless, attempts to address these issues through the publication of AI ethics and governance frameworks have been proliferating in recent years, with governments and technology companies producing hundreds of sets of ethical codes and guidelines. At the international level alone, the OECD, G20, GPAI, the Council of Europe, the European Union and UNESCO have all been working on projects looking at the ethics and governance of AI. The European Commission recently published its draft regulation on AI, which proposed a risk-based approach to governing the technology, and the Council of Europe is in the process of exploring the feasibility of an international legal framework to ensure that the design, development, and deployment of AI systems accord with the principles of human rights, the functioning of democracy, and the observance of the rule of law. In November 2021, UNESCO published and adopted the first global ethics framework (UNESCO 2021).

But in this international rush to find acceptable and lasting global standards for developing and using AI, who is being left out of the discussion? Is it merely "ethics washing", as many critics assert? Whose voices are being listened to, whose values are ultimately being represented in these ethical codes and governance frameworks, and how are these codes and frameworks being followed or enforced in actual practice? Many of these efforts, even at the international level, are centred on small groups of experts in Europe and North America. This is a problem given the truly planetary sweep of the technology and the reverberating effects of its applications. As Kate Crawford writes in her 2021 book, *The Atlas of AI*, AI infrastructure is as "hybrid" as it is transnational: dependent on the mining of natural resources, human labour, and data from around the world.

PATH-AI

What might the ethics and governance of AI look like if they were decentred from "Western"- dominated perspectives, and instead incorporated approaches, ideas, values, and wisdom from other places? How are values central to Anglo-European and American discourses around the development and use of AI systems, such as privacy, agency and trust, viewed in other cultures? How can different understandings of these values be squared with the international ethics and governance regimes centred on Anglo-European and American philosophical underpinnings that we have seen emerging around AI? And how can these different understandings help to set the direction of travel for responsible and sustainable AI innovation in the society of tomorrow?

These are some of the questions we have been grappling with through PATH-AI. PATH-AI (Privacy, Agency and Trust in Human-AI Ecosystems) is a collaborative research project between the UK and Japan, involving The Alan Turing Institute, the University of Edinburgh, and RIKEN, Japan's largest comprehensive research institution. The aim of this project is to understand how the three interrelated values of privacy, agency and trust work together in very different socio-cultural contexts in relation to AI and other data-driven technologies. Although we may think of such values as more or less universal, this is not the case: in fact, as we will see, there is considerable variation in their meaning and context of application. Our research so far has shown that this variation derives, at least in part, from differing social and political structures and institutional configurations, philosophical traditions, and underlying assumptions about personhood and its relationship to technology.

Too often, the most globally influential debates about AI ethics and governance have resembled a top-down exercise conducted by a small group of experts educated in Western Europe or North America. PATH-AI aims to start recognising and redressing this bias, in part by gathering a diverse range of views. We conducted interviews and questionnaires involving 95 people, split between the UK and Japan, and between experts and non-experts. We aim to take an approach that incorporates, and learns from, people's everyday ethical values, concerns, beliefs and lived experience.

PATH-AI began in January 2020 against the backdrop of the unfolding COVID-19 pandemic, and has of course been influenced by these events in that we shifted much of the focus of the research to look at British and Japanese governmental responses to the pandemic in terms of our main concepts. In our interviews and questionnaires, conducted between April and August 2021, we asked people about their views on and use of three emerging data-driven healthcare technologies: digital contact tracing apps, symptom checking tools, and care robots.

Digital contact tracing apps included the NHS COVID-19 app for England and Wales and Japan's national COCOA app. Medical symptom checking tools included websites such as the NHS' online 111 service as well as AI-driven symptom checking services from private providers. Care robots were defined for the purposes of the study as physical robots that could hold simple conversations and might be used for tasks such as keeping older people company, telling jokes, or reminding users to take their medicine. We also asked questions designed to draw out their views on privacy, agency, and trust in relation to these and other AI and data-driven technologies.

The pandemic has served as a major test of these values, in both the UK and Japan. People's *agency* has been curtailed in lockdowns and states of emergency, both of which have placed previously unthinkable legal restrictions on liberties; governments have asked for and called upon *trust* in enacting these measures, in tracking and tracing citizens, and in undertaking new forms of public health surveillance; the implementation of technologies such as digital contact tracing apps and debates about "immunity passports" have also brought concerns of *privacy* to the fore. Data, in many different forms, has become a vital need in understanding and managing the pandemic, but also a battleground in the fight against false or misleading information (the "infodemic"), and a difficult-to-calibrate tool for controlling the spread of the infection (as seen with the "pingdemic"). It has also, in the UK context at least, become a kind of prism through which the white light of digital privilege and power has been refracted – exposing how systemic inequalities, structural injustices, and digital divides are manifest in imbalanced and unrepresentative datasets (for example, see Public Health England 2020).

By gathering views about these issues and the values which might inform and transform them, we aim to draw out differences and similarities between the UK and Japanese cases, and develop an intercultural perspective upon which to construct a new approach to governance frameworks. A full interim report will be published in early 2022, but this short-form version is intended to present a brief outline of some of our initial findings. The following sections provide short overviews of the concepts of privacy, agency, and trust in the UK and Japanese contexts as they relate to AI and data-driven technologies, a summary of the findings of the interviews and questionnaires, and a brief conclusion setting out some key themes that we will be pursuing in the next stage of the project.

Privacy

As the development and implementation of data-intensive digital technologies have accelerated over the past decade, debates about data protection, confidentiality, and privacy in the age of "surveillance capitalism" (Zuboff 2019) have also moved to the fore. Instances of data leaks and unlawful data sharing have grown steadily in scale and severity: most infamously in the form of Edward Snowden's revelations about mass government surveillance, and the exposure of Cambridge Analytica's use of data collected via Facebook without user consent to micro-target political adverts (Cadwalladr and Graham-Harrison 2018). In the UK public sector we have also seen, for example, the sharing of NHS data with Alphabet company DeepMind and the use of live facial recognition by South Wales Police (the case of Bridges v SWP 2020) – both breaches of the Data Protection Act (ICO 2017, ICO 2019), and in the latter case, also a breach of the Equality Act and the European Convention on Human Rights.[1] At the same time, lawful forms of mass data collection by corporations have also continued to intensify.

The COVID-19 pandemic has served as a catalyst for these issues. Digital technologies have come to permeate many people's daily lives, education, and work to a greater extent than ever before, and are continuing to redefine our relationships with government, businesses, and each other. Tech corporations together with governments have implemented various kinds of test, track and trace infrastructures, as well as vaccination passports or immunity certificates, which have been presented as ways out of the pandemic. New surveillance tools have been introduced for companies, universities, and schools to keep tabs on employees and students working from home. People are being digitally "tracked" and "traced" by a variety of actors in their daily lives to an unprecedented degree, and this looks set to continue beyond the current health emergency.

As increasingly complex and wider-ranging algorithmic systems are being introduced, as more public services are being digitised, and with a handful of large tech companies claiming near-monopolies over the vast swathes of consumer data being generated in particular domains, public concerns about potential violations of privacy rights and about the individual- and community-level impacts of intrusive data collection and use appear to be growing. Many people choose a more privacy-preserving approach when given the option. For example, millions of users abandoned WhatsApp for rival messaging app Signal after the former updated its privacy policy in 2021 to allow the sharing of some data with its parent company Facebook (Hern 2021). Later in the same year, when an Apple operating system update gave users the option to prevent apps tracking their digital activity across other apps and websites, almost all users in the US opted out of sharing their data in this way (Bensinger 2021).

1 This case did not, however, result in an outright ban of the use of live facial recognition by police (see Bridges vs SWP 2020).

Perhaps, then, it is unsurprising that privacy has come to be seen by some as a "hegemonic value"[2] – *the* central focus of the majority of government, corporate, media, and public attention in relation to data-intensive and AI technologies, particularly in Europe. Although "data protection" and "privacy" are not, strictly speaking, interchangeable terms, new legislation in the UK and Europe has centred around data protection, most notably via the EU's General Data Protection Regulation (GDPR), which was incorporated into UK law as the Data Protection Act 2018 and introduced new, more stringent rules around the use of data.[3] The same focus was seen during the development of digital contact tracing apps, when early attempts by several governments, including the UK and Japanese governments, to develop apps that would store user data on a central database accessible to the state for analytical purposes were generally overturned in favour of Google and Apple's "privacy-preserving" decentralised approach. The Google/Apple app framework allowed the storage of only anonymised data on individual users' devices and relied on Bluetooth signals instead of GPS data to ensure that the user's location and contacts could not be identified. While this ostensibly safeguarded users' privacy, it also effectively privatised a critical corner of public health infrastructure, hampering public health officials' capacity to use data for the public benefit and setting the terms for the provision of safety-critical social services from Silicon Valley rather than from local centres of healthcare practice where community-connected understandings may have otherwise shaped public health strategies and agendas.

The story of the value of privacy in the contemporary UK context is thus a complex one. On the one hand, there seems to be growing public awareness that consumers pay for access to "free" social media and online services with their data, and that the potential profit-making uses of this data by companies are far-reaching and could have both positive and negative direct and indirect consequences for individuals, communities, and societies both now and into the future. On the other, there is now more legislation in place in the UK and Europe focusing on protection of personal data as well as greater emphasis on privacy from some tech companies. Yet this focus on individual privacy rights often is claimed to override the use of data for public wellbeing, and to curtail governments' powers over online spaces and activities, as in the case of end-to-end encryption of private messaging applications. A number of questions and concerns persist.

The first is whether consumers' understanding of issues around privacy and data protection is keeping pace with increasingly complex legislation and regulations, as well as with how data is actually being used by companies. This is particularly the case in opaque and proprietary areas such as machine learning that are not accessible or explained to those who are subjected to them and who may not even be aware that they are having their data processed by these technologies. In an age in which impoverished forms of "informed consent" largely consist of a deluge of cookie notices and unread terms and conditions statements, the "responsibilisation" of individual consumers to gatekeep the extractive objectives of tech companies signals a deeper societal problem that is in dire need of attention.

2 This phrase is taken from a TILTing 2021 conference panel entitled "Privacy's Value Hegemony in the Context of Technological Responses to the Pandemic", with contributions from Tamar Sharon, Marjolein Lanzing, Lotje Siffels, Natali Helberger, Sarah Eskens, Joanna Strycharz, Joran van Apeldoorn, and Marijn Sax.

3 In August and September 2021, the UK government announced plans to diverge from parts of the GDPR perceived as being "disproportionate" and overly onerous (UK Government 2021).

A second is the danger of conflating privacy with the far narrower realms of data protection or confidentiality. Data protection is a subset of information privacy, which in turn is a subset of privacy, while confidentiality focuses on rules or promises that limit access to certain types of information. The GDPR covers data protection rather than privacy more broadly. Moreover, there is a question of whether protection of identity *per se* is quite as central a value as it used to be, since powerful edge computing technologies may mean that a company no longer needs to link a user's data to their real-world identity in order to micro-profile and target them, and to attempt to predict or manipulate their behaviour. Finally, there is the question of whether problems around privacy that arise from the application of new technologies like AI are best addressed through more technology, new legislation and regulations, the development and enforcement of existing regulatory and legal apparatuses, or other alternatives such as shifts in societal norms and heightened public awareness. A combination of these and other instruments, which can be mutually supportive, is likely to be a more robust approach to protecting privacy than reliance on tools used singly (Bennett and Raab 2006).

Privacy is a key issue in Japan as it is in the UK. Although the Japanese word used to translate privacy (*puraibashii*) was only formally introduced in the 1960s, indigenous conceptions of privacy predate this (Adams *et al.* 2009, Hildebrandt 2015). But what the concept of privacy consisted of was and remains different from what the term means in the UK. American and Anglo-European ideas of privacy have tended to be understood in terms of "the right to be let alone" (Warren and Brandeis 1890) and having individual control of one's personal data (Westin 1967). This is broadly premised on the European liberal philosophical tradition of the autonomous and "self-possessing" individual, who has rights and prerogatives to pursue their personal self-interest above all else. This is reflected, for example, in the GDPR, which assigns legal rights to citizens as individuals. However, recent work has recognised that people do not always think of privacy this way in their daily lives, and there has been growing interest in understanding privacy *relationally* and as a social and political good that bears upon the wellbeing of collective life, including democracy. As Bannerman puts it, "Relational thinking shifts privacy regulation from a focus on withdrawal or the sheltering of particular communications, to an emphasis on how networked relations foster, or fail to foster, healthy communities and relationally autonomous selves" (Bannerman 2018, 8; see also Raab 2012).

In Japan, the relational sense of privacy has tended to predominate, with privacy understood as context-dependent and reliant on the behaviour and context of the environment and other people, carefully calibrated to social situation and the specific person, organisation or entity one is interacting with rather than being based on an absolute individual right that can be asserted and backed up by law. In this sense, as Hildebrandt puts it, "The question is not whether a person 'has' privacy, but whether the environment 'affords' a person a degree of privacy" (2015, 104). This in turn is very similar to some American and Anglo-European approaches, such as Nissenbaum's (2010) focus on contextual privacy. But values change: Japan has now seen three decades of neoliberal state efforts to bolster the cultural ideal of the individual, with greater focus on individual rights and choices. Arguably one sign – and catalyst – of this shift has been the passing of Japan's Act on the Protection of Personal Information (APPI) in 2003, which received "adequacy" status with GDPR in January 2019, and moved the country further towards the European approach to data protection.

Agency

"It's like Big Brother following you. You feel that you're being watched, you can't go anywhere, they're watching what you're doing. You just want some privacy, and your choice is being taken away from you, you want to be able to choose what you do, where you go and just get on with life on your own terms."

(UK general public interviewee)

This response came from an interviewee concerned about the NHS COVID-19 digital contact tracing app released in 2020, and points up the close relationship between privacy and agency: one facilitates the other. Although the app itself only collects anonymised data and stores it on the user's device rather than in a centralised database controlled by the NHS or government, several interviewees nevertheless worried that it was simply the latest and most blatant instance of a broader wave of digital technologies tracking their daily activities – infringing on their privacy and constraining their sense of agency.

By agency, we mean in Anglo-European and American contexts the ability or potential to act on one's own volition, to make one's own choices about how to live and flourish, and to freely pursue one's own life path. As noted above, liberalism is premised on the concepts of individualism and independence. But agency itself is a complex concept that, like privacy, is not immediately reducible to the autonomous and self-possessing individual. The formative dimension of a person's identity is always already constituted in social relations and demands for integration into societal (as well as political and economic) norms and institutions. From the earliest stages of the child-caregiver relationship to the modes of mutual recognition through which adults gain a sense of respect and self-esteem, processes of identity formation are entwined with interpersonal communication and relationships of social cohesion and reciprocal trust. Privacy and other "individual" human rights, of course, protect these activities and relationships. Along these lines, the decisions we make and the life trajectories we choose are never really unconstrained or independent in a strict sense but are rather relationally implicated. Agency, like privacy, can be understood relationally: as emerging from a field of relations rather than from an individual's will or capacities – and these relations can include technologies. Again, this goes to fundamental questions about how we understand our selves and our inner lives in relation to others in society.

Models of how the self is understood in Japan have tended to foreground this relationality and interdependence with others, as for example in Doi's (1973) influential though much-critiqued work on *amae* (loosely translated as "dependence") as a foundational and ideal component of Japanese social relationships.

The self has also often been presented as fluid – dynamically adjusted to the particular social context and one's relationship to people deemed insiders or outsiders within this context (Adams et al. 2009), even while strong norms exist that prescribe specific and clearly delineated modes of behaviour within certain contexts and roles – to a far greater extent than in Anglo-European and American societies. There is no exactly comparable word in Japanese to translate "agency."[4] However, as with the word *puraibashii* ("privacy"), which only formally entered the Japanese language in the 1960s, the lack of a directly equivalent word does not mean that similar concepts do not exist in other forms. In this case it may indicate other concerns, however, such as a greater importance placed on group rather than individual or independent action, capabilities, or decision-making. Adams et al. (2009) characterise the way in which Japanese society is made up of groups with strong vertical but weak horizontal ties as "insular collectivism", whereby the needs and interests of these groups have – historically, at least – come ahead of those of individuals, with a fragmentation of society into competing organisational "islands".

What does this mean in relation to AI? AI systems can be developed, implemented and controlled in ways that constrain or impact individual, relational or collective agency in a variety of ways. AI-enabled surveillance technologies and data analytics can affect how people act when they know they are being watched or monitored, with a potential chilling effect on agency-related behaviour, including social interaction, physical movement, the use of public space, and communications. AI is used to determine what search engine results, news articles, social media posts or YouTube videos are presented or recommended to us as individuals and groups, and can thus potentially manipulate the frames of reference and sources of information based on which we might make a particular decision. AI could also be used in the form of automated decision-making algorithms that may, for example, reject a person's application for a loan based on a variety of factors to which they may not have access or that may not be fully explained to them. These same decision subjects may then have very limited opportunity to appeal the decision. In some cases, people may not be aware of other decision-making algorithms that influence their lives, for example if a CV is automatically rejected by an algorithm before even reaching the desk of a flesh-and-blood recruitment manager. Resultant failures to secure a job or a loan are likely to have a significant impact on one's ability to exercise agency. In theory, the latter examples of automated decision-making without the subject's being informed of the rationale for the decision should no longer happen in the UK or EU under the GDPR's provision of a "right to an explanation" (Article 22),[5] although its status as a right, its meaning, and the likelihood of its being satisfactorily implemented in practice have all been disputed. AI systems trained on datasets that originate in human actions, relations, and institutions can also reinforce bias, baking in structural forms of discrimination and solidifying constraints on agency: quite literally, in the case of a parole risk assessment algorithm in the US that was found to provide recommendations for incarceration that discriminated against black men (Angwin et al. 2016). In ways we may not even be aware of, AI could reframe the information available to us for making decisions, nudge and influence those decisions, pre-empt the presentation of certain decision alternatives, or even make some of them for us.

4 "Agency" in the context of social science and philosophy tends to be translated as *shutaisei* ("subjectivity" or "autonomy") or *kōishutaisei* (literally "action subjectivity" or "action autonomy"), or simply transliterated as *ējenshī*.
5 Article 22 requires that any company conducting automated decision-making must give individuals information about the processing and provide simple ways for them to request human intervention or challenge a decision (see https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling/).

Of course, this is not to say that similar kinds of societal constraint did not exist before AI came along, also in more or less visible manifestations. Subtle and opaque forms of ideological and discursive control predate the digital age, and methods and mechanisms of agenda-setting and agency-shaping power have operated under the skin of social institutions since time immemorial. However, in the process of translating these existing patterns and mechanisms of power and control into digital processes and systems, the development and application of AI technologies –enjoying the prestige and aura of infallibility that "technology" has in many societies – are forcing us nevertheless to confront ugly histories of structural violence, systemic inequality, and discrimination based on factors such as race, gender, class, disability, and sexuality. In this sense, the current moment could present an opportunity to build more equitable societies, at the same time as it carries the troubling risk of deepening these same de-agentifying forms of power and control.

To complicate this narrative about agency, even while humans may have their own agency constrained, AI systems are themselves increasingly being given, in the minds of some, a form of distributed agency – an ability to act with some degree of perceived autonomy. This is clear in the term "AI agent", used to refer to an AI application that acts on a person's behalf in some way. For example, the Institute of Electrical and Electronics Engineers (IEEE) has proposed the creation of an AI agent envisaged to extend the user's agency by acting as a kind of personal privacy lawyer-cum-sales agent, making deals with the user's personal data according to their pre-set wishes and own customised terms and conditions (IEEE 2019). This seems to imagine a future in which such negotiations have become a moment-by-moment aspect and condition of moving through the world rather than simply a response to a particular data request. In another direction, the advent of AI-assisted "smart" devices promises to enhance our own agency through the liberating removal of burdensome chores and inconveniences that may detract from our ability to develop our selves, or to exercise choice in lifestyles and relationships – which presents something of a paradox, since we are also expected to delegate agency to these technologies.

Some have suggested that Japan's historical mixture of Buddhist, Shinto and Confucian religious and cultural influences means Japanese people have a greater capacity and willingness to relate to technological objects or devices, such as robots or AI agents, as companions or colleagues imbued with their own spirit, rather than as straightforward tools—as they are more commonly seen in the UK (Hildebrandt 2015, Gal 2020). Others have disputed this as a fundamental misunderstanding of these cultural influences (Frumer 2018, Gygi 2018). Instead, the relentless promotion by domestic media and politicians of a narrative portraying Japan as a futuristic nation of technophiles may play a more significant role. Either way, ultimately, the idea of an AI system having agency seems itself to be problematic, as it disguises the fact that such systems are human creations and implies that human actors may no longer be held accountable for the behaviour and consequences of the tools they build. Nevertheless, it is important to understand the extent to which people in both the UK and Japan see their actions and capabilities as constrained or extended by AI and related digital technologies.

It is clear that, as the semantics of the term “agency” evolve with the development of AI technologies, there will be a need to continually re-assess how we use this concept. The term “agency” has a multi-layered character with meanings that range across practical, functional, and normative levels and that must be explicated with keen awareness of the subtle differences between these valences. The stakes of such interpretive understandings reach into the core of how we might define the humanity of the human in an age of rapidly growing digitisation, and how we might understand future possibilities for the actualisation of freedom.

Trust

Public trust in the use of digital technologies became a key issue in the government responses to the COVID-19 pandemic in both the UK and Japan. The roll-out of the NHS COVID-19 app, and users' adherence to the request to self-isolate if pinged, were initially presented by the UK government in 2020 as essential determinants of the success or failure of the wider NHS Test and Trace system. Studies suggested that the app would only be substantially effective in preventing the spread of the virus if a relatively high percentage of the population downloaded and used it (University of Oxford 2020): if people did not trust the government, or if they did not trust that others would download and activate it, they might not do so themselves. As a report by the Ada Lovelace Institute on COVID-19 technologies, including digital contact tracing apps, noted: "Public trust is essential for any technological system that is deployed widely and with significant impacts across society" (Ada Lovelace Institute, 2020) – although interestingly, in our interviews with UK research participants, we found that a sense of downloading the NHS COVID-19 app for the greater good generally overrode lack of trust in the government. In Japan, the level of trust in government is relatively low by international standards (Edelman 2021). In order to encourage compliance with COVID restrictions despite widespread public distrust, politicians in Japan leveraged the concept of "self-restraint" (*jishuku*), which has become linked to a sense of Japanese cultural identity, while at the same time tolerating or even in some cases actively engaging in the public shaming of individuals and businesses who failed to comply with government guidelines (Wright 2021).

Sociologists and political scientists divide trust up into various categories, including public trust in governments and politicians, trust in institutions, and interpersonal trust between fellow citizens. Political trust may be undermined by scandals or bolstered by transparency and clear communication. Interpersonal trust forms an implicit pillar of social order in the modern era. When individual human beings behave and act in ways that affect one another for better or worse, contemporary society binds them to the justifiability of their actions based upon reasonable expectations that they will exercise good judgment in pursuing their objectives in ways that do not harm those around them. Indeed, securing such a connection between responsibility and interpersonal trust involves establishing a stable bedrock of behavioural expectations between rational agents whereby mutually accountable performances can be universally assumed (Bauer and Freitag 2018). Such a stable starting point for free and orderly social coexistence has been variously called "basic trust" (Erikson 1959), "generalised trust" (Uslaner 2002), and "generalised expectancy" (Rotter 1967). Trust, writ larger, is consequently seen as a "good thing": a vital component of social capital, or the stuff that emotively and rationally binds people together, allows bridges to form between individuals and groups who might have divergent beliefs or interests, and enables society as a whole to stick together and prosper.

But what, exactly, people or institutions are trusted to do or not do, be or not be, and so on, is important in our understanding of how trust functions. In other words, the moral or functional qualities that attract our trust in politicians, technologies, professionals, scientists, the media and other objects and roles are part of the trust dimension. This is because they are the criteria by which we assess these objects' trustworthiness. How we make such judgments, and the factors that affect them, are important. Their protectiveness, fostering or denial of our privacy or agency might be among the available criteria for trustworthiness evaluations.

Trust also requires uncertainty and freedom of action for the person being trusted – otherwise it is not trust but control: a restriction of agency in the service of making behaviour reliably predictable. Japan is a particularly interesting example of this tension. On the one hand, it is fairly common practice to leave one's wallet or smartphone unattended in a crowded bar in the centre of Tokyo to hold a table without fear of it being stolen; Francis Fukuyama (1996) has argued that Japan's economic success is partly based on the high degree of trust in other people and reliable expectations about how they will behave. On the other hand, by focusing on the need for freedom of action in defining trust, Toshio Yamagishi has argued that Japan is a low-trust culture. He argues that this is because "the collectivist society produces security but destroys trust" (Yamagishi 2011, 1). In other words, if you do not have true choice of action because of tight societal constraints, it is not genuine trust. Yamagishi links the possibility of trust to individualism – a view closer to the liberal idea of the individual self that has the capacity to make more independent decisions.

Again, it is important to note that in the decades since the bursting of Japan's bubble economy at the end of the 1980s, neoliberal policies from the ruling Liberal Democratic Party administration together with changing cultural influences have often focused on promoting individualistic aspirations and freedom of choice (Lukács 2010), while economic inequality has risen steadily together with forms of non-regular employment. Public discourse has vacillated between ongoing affirmation of a shared communitarian ethno-national Japanese identity and an alternative view of Japan as a society where the ties that bind have been severed (*muen shakai*), precarity and loneliness have become normalised (Allison 2013), and levels of trust have declined (Lukner and Sakaki 2017).

Here we come back to the potential impacts of particular applications of AI. If AI systems constrain our agency in decision-making, displace human actions that would otherwise strengthen mutual expectations of reasonable behaviour, and exacerbate forms of social and economic inequality, there is a danger of expectations of trust giving way to control and to the deterioration of the bonds of reciprocal responsibility: less a case of "trust, but verify" and more a case of "never trust, always verify". This can be seen as both a digitalisation and expansion of existing and increasingly ubiquitous audit culture – the tendency for public institutions and companies alike to perform responsibility to stakeholders by continuously checking everyone's actions (Power 1997). It can also be seen as an extension of the atomising logic of neoliberal modes of responsibilisation.

Counter-intuitively, shifting the onus of mutual accountability entirely to the individual level maintains vertical forms of socioeconomic hegemony while undermining the horizontal forms of societal reciprocity that have preserved both interpersonal and institutional bonds of trust in the modern democratic age. Along similar lines, the shift from trust to control made possible and indeed encouraged by AI-driven automation can be seen as an outgrowth of a prison-like logic of surveillance facilitated by AI technologies like facial recognition systems—a shift that started with the security services before moving into every other aspect of life, from policing to education to healthcare (Kuldova 2020, Crawford 2021).

How can we protect space for trust and agency in a society of omnipresent "algorithmic governance", in which citizens and decision subjects are compelled to rely upon algorithms instead of, for example, elected representatives or company executives, to choose the right course of action and determine the extent and reach of the choices that are available to them? How, and exactly where, should human agents be "in the loop" in automation-assisted decision-making systems? To what degree should we be restricting how AI is used when it comes to these complex and broad-reaching issues? Should agency itself be understood less in terms of the individual and more in terms of the social relation? Can AI tools be regulated in such a way as to multiply rather than reduce the agency of individuals, communities, and societies?

Overview of research findings

"It is necessary to open up the contents of AI technology, have experts discuss problems and issues, and make the contents public. The situation at the moment is that it is done in a closed room, so the general public does not know."

(Japanese general public respondent)

To think through some of these ideas from a more bottom-up perspective and understand how they relate to people's lived experience and disparate views about and uses of digital technologies, the PATH-AI teams in the UK and Japan conducted semi-structured interviews and questionnaires with members of the general public as well as experts. In the UK, interviews were conducted with 26 members of the public who responded to a request for research participants kindly publicised by Camden Council and Carers UK. We also approached and interviewed 17 experts drawn from across the public sector, business, academia, and third sector, who were working on AI and/or digital healthcare technology. In Japan, a survey was conducted by Intage, a marketing research company, on behalf of the PATH-AI Japan team. 501 members of the general public were surveyed, of which a sample of 26 answer sets was provided to the UK PATH-AI team for analysis. A slightly more detailed survey was sent to and completed by 23 experts, and a further 3 experts were interviewed. A similar question set (translated from English into Japanese) was used as the basis for all of the interviews and questionnaires, which took place between April and August 2021.

All research participants were asked about three different examples of data-intensive digital technologies in the areas of health and social care: digital contact tracing apps, medical symptom checking tools, and care robots. As previously mentioned, digital contact tracing apps included the NHS COVID-19 app for England and Wales, the Protect Scotland app, and Japan's national COCOA app. Medical symptom checking tools included websites such as the NHS' online 111 service as well as AI-driven symptom checking services from private providers. Care robots were defined for the purposes of the study as physical robots that could hold simple conversations and might be used for tasks such as keeping older people company, telling jokes, or reminding users to take their medicine.

"Part of the problem with so many of these international principles, like GPAI and OECD is that they are extremely Western-centric"

(UK expert interviewee)

The decision to focus on these case studies was driven in part by the COVID-19 pandemic, during which such tools gained prominence as ways to reduce the spread of infection and cope with the shifting demands placed on healthcare systems. These are technologies that seem to offer huge potential benefits, but also raise many questions about ethics and governance. Although they serve different functions, what they have in common is their digital and data-driven nature and their promise of revolutionising health and care systems by transforming the very infrastructure of healthcare. As we will see, similar concerns were expressed by research participants about the storage and use of data collected by these tools, who is benefiting or should benefit from their use, and what form regulation and governance should take in order to ensure that they benefit the public good. Research participants were also asked for their views on privacy in relation to the sharing of their healthcare data, their level of trust in government and technology companies and how this affected their use of such tools, how they viewed the role of ordinary citizens in how these technologies are regulated and governed, as well as other questions designed to elicit their views on the values of privacy, agency, and trust.

It is important to note the limitations of this approach. It was only possible to conduct semi-structured interviews – which provided richer qualitative data than the questionnaire responses – in the UK due to the inability of UK researchers to travel to Japan because of the pandemic combined with resource limitations in Japan. The samples were also not randomly selected and not large enough to be representative of particular expert communities let alone the broader general public in these two countries; the findings cannot be generalised to any wider populations. The intention was rather to gather a variety of different perspectives and insights on these issues, and to consider a wide range of views while also identifying broad areas of consensus and divergence that could inform the next stage of our research project.

A more comprehensive version of our findings, which lays out the research methodology and results in more detail, will be published in the full report in early 2022. Here we present a very brief overview organised by topic.

"If you tie your own hands when it comes to AI, you will definitely end up behind other countries. I want a compromise to be found between flexibility and responsibility"

(Japanese general public respondent)

Digital contact tracing apps

- UK interviewees' accounts of the central government's handling of the pandemic, including the development of the NHS COVID-19 app as part of the NHS Test and Trace system, were dominated by strongly negative views of politicians and senior officials, mistrust of the central government, and perceptions of their incompetent handling of aspects of the pandemic response. Many interviewees stated that the government had ignored the views of citizens and experts they disagreed with.
- Both UK and Japanese respondents suggested that the early choice by both national governments to develop a digital contact tracing (DCT) app with a centralised database contributed to privacy concerns and mistrust that persisted even after this strategy was abandoned and a privacy-friendly decentralised app using the Google/Apple framework was developed in its place.
- While UK and Japanese respondents largely praised the *concept* of DCT, they also critiqued the reality as almost entirely ineffective, in part because of the decision to use a decentralised app employing Bluetooth to detect contacts. Experts in both countries talked about a trade-off between privacy and efficacy in deciding between the centralised and decentralised approaches, and complained about the absence of substantive public debate or consultation about the nature of this trade-off, as well as the lack of integration of the apps that were eventually developed into the broader pandemic strategy and healthcare system.
- In relation to their decisions about whether to download and use a DCT app, several Japanese respondents explicitly mentioned a sense of patriotic, civil or societal duty, as well as greater concern about what others around them were doing, which seemed to exert a strong influence on their own actions. UK interviewees tended to talk in slightly different terms of social responsibility and "doing the right thing".

> "I didn't particularly trust the government, I just saw [using the NHS COVID-19 app] as a necessary thing like a lot of the necessary things that we have to do. I call them hoops that you have to jump through to carry on with your daily life at the moment."
>
> (UK general public interviewee)

- UK interviewees suggested several lessons to be learned to rebuild trust and do better in the future. These included that the government should better consult with, listen to, and communicate with citizens, and that citizens themselves should be empowered with clear information and digital tools that would enable them to exercise greater agency instead of being presented with black-box solutions like DCT apps that provided them with little information or context and therefore little sense of agency.
- Lessons from Japanese respondents focused on competence, including the need for the government to overcome bureaucratic siloes, and to learn lessons on more effective IT design. Several Japanese respondents suggested that making the app mandatory – making its use a "duty" and not a personal choice – would also have been a more effective strategy.

> "Because [COCOA, the Japanese equivalent of the NHS COVID-19 app] has bugs, and therefore I can't trust it as an app, I don't want to download it."
>
> (Japanese general public respondent)

Symptom checking tools

- All groups of respondents in both Japan and the UK raised concerns about the effectiveness and accuracy of symptom checkers, and whether possible harms – such as under- or over-diagnosis, or feeding anxiety or hypochondria among users – outweighed the possible benefits of reducing the burden on healthcare systems, democratising medical expertise, providing a more convenient and user-friendly experience, and perhaps ultimately even improving the quality of healthcare.
- Respondents suggested that the current range of symptom checking tools available had both benefits and drawbacks, and that they were likely to be most useful in providing basic reassurance over very minor ailments. However, most were in agreement that none of these tools were yet close to replicating the benefit of seeing a healthcare professional face-to-face.
- UK experts argued that the UK regulatory environment was relatively permissive, with symptom checking services seemingly operating in a grey area and technology companies taking advantage of this by rolling out new AI-powered tools without strong clinical evidence and with heavy legal caveating in their small print, in a manner that may breach existing regulations. By contrast, in Japan, the lack of clear regulation together with the concentration of power with doctors in Japan's health system had the opposite effect, restricting companies in this space so there were very few services available and those that existed seemed more limited in scope compared with counterparts operating in the UK.
- UK concerns especially among experts included the idea that tech companies could use symptom checkers as a back door to privatise parts of the NHS by accumulating large amounts of people's medical and biometric data. This was exacerbated by the growing difficulty experienced by many interviewees in accessing GP services even before the pandemic and a suspicion that symptom checkers could evolve into a gatekeeping tool to restrict access to or even replace in-person services.[6] In Japan, where healthcare is regulated and funded by a national health insurance system but privately provided, the same kind of concern did not exist.

> "People talk about data as oil. Somebody else said, no, it's not. It's blood. It's given by patients. We should be exercising some control over how it's used."
>
> (UK expert interviewee)

6 In fact, the former had indeed happened in the case of NHS COVID-19 symptom checking tools (Mansab et al. 2021).

- In both contexts, however, there was considerable distrust of multinational tech companies becoming increasingly involved in healthcare, and worries about how data gathered by symptom checking and other digital health tracking tools might be used or monetised.
- Experts suggested the need for far greater transparency and clarity, as well as clearer articulation and enforcement of existing regulation with the possibility for new sectoral regulation if necessary. They also expressed concerns about the fragmentation of data across different digital systems, and called for ensuring the portability and shareability of data and standardisation of data formats.

> "[Privacy protection] should be decided by the balance between the benefits received from the AI system and its risks and costs."
>
> (Japanese expert respondent)

Care robots

- The interviews and surveys revealed particularly striking differences in attitudes towards care robots.
- UK views among both the general public and experts were highly polarised, and largely dominated by what one interviewee called “the ick factor” – a sense of disquiet or even disgust at the idea of (particularly humanoid) robots caring for older adults at the expense of human care. Several interviewees suggested the use of care robots would be a sad reflection of a society that did not value care givers or older people.
- UK interviewees raised an array of concerns, including about: the potential for dehumanisation of care; the quality of social interactions with robots compared to those with humans; the functional abilities of robots; the cost of purchasing robots and how investment in them could divert funds from human care givers; potential physical risks to care receivers; privacy concerns; and the potential for older adults to become emotionally attached to their robots in a problematic manner. Some positive views were also expressed, particularly about the potential of more functional robots or pet-type companion robots.
- There was a general consensus that such robots would probably be increasingly used, primarily in order to reduce the cost of care. Expert interviewees were more open than general public interviewees to the idea of using care robots, but argued that more evidence was needed about their practical efficacy and costs and benefits.
- Most UK interviewees concluded that robots did “have a place”, but many were unclear about what that place should be, and what exactly these devices should look like and do. However, there was complete agreement that care robots should only ever be used as a supplement rather than a substitute for human care.
- In Japan, by contrast, attitudes towards care robots were far more positive. No “ick factor” was apparent as around 90% of respondents answered that care robots were a good idea. Discussion of their presumed benefits largely centred around helping address the labour shortage in the care sector and relieving the burden on human care givers, as well as some comments in support of robots replacing care workers altogether.

- Several Japanese respondents noted that care robots could provide older people with useful daily interaction that they might otherwise lack, that having a care robot was better than being left alone, and that robots could help older people, who might otherwise have to rely on a human care giver, to retain their sense of self-esteem.
- Concerns in Japan centred instead on the cyber security of the robots and the possibility that they could be hacked and used to manipulate or defraud vulnerable users, as well as on the need to elaborate where legal responsibility lay in the result of a user being harmed by a robot.

> "We need to be able to put in the hands of citizens AI and machine learning technologies which work on their behalf. And that should not be Siri or Google."
>
> (UK expert interviewee)

Privacy, agency and trust

"[Privacy is] a losing battle! I have no hope of being private, I mean the reality is that, even if you are the most privacy conscious person in the world, good luck, no way."

(UK expert interviewee)

- Many UK respondents expressed qualified trust for the NHS brand despite the perceived growing encroachment of politics, but also said they did not trust the organisation's technological competence. The converse was true for tech companies: their technological competence and controls were highly trusted, but most respondents did not expect them to act in anyone's interest but their own. Many said they had lost trust in the government over several high-profile scandals during the COVID pandemic.
- Japanese respondents answered that they had relatively little trust in tech companies but even less in government, while the majority agreed that public trust was an important condition for the widespread deployment of AI.
- Many UK interviewees discussed the recent failed NHS data sharing plan (General Practice Data for Planning and Research, or GPDPR) as an example of how such data should *not* be shared. Several expert interviewees argued that while they completely agreed with the idea behind GPDPR, its communication had been extremely poor and this together with its misrepresentation in the media had undermined public trust and fuelled people's suspicions about the NHS selling off their data. The majority of UK respondents said that they wanted clearer information about and control over who used their health and care data, how, and for what purpose.
- These concerns were echoed by many Japanese respondents, who repeatedly emphasised the importance of privacy and concerns about potential data leakage.
- However, several UK and Japanese experts argued that while privacy and data protection were important, "privacy fundamentalism" was in danger of jeopardising the huge potential benefits of sharing healthcare data. Some argued for an approach that put more tools in the hands of citizens and empowered them to exercise greater agency over their own data, while others proposed a more collective way forward that did not focus on individual privacy but instead on public outcomes.

"I would want to know where [my personal data] was going, and probably what it would be used for... also how long the data would be held for."

(UK general public interviewee)

- Across both the UK and Japan, experts were split on how to regulate AI and the data sharing that would power it. Most argued that the EU's proposed regulation of AI was too cumbersome and its attempt to regulate AI as a general technology the wrong strategy, proposing instead a sector-by-sector approach. There was also widespread agreement on the need to embrace a mixture of measures that included laws, regulations, ethical principles and education of developers and users.
- Both general public and expert interviewees said that it was important for ordinary citizens to have a substantive say in the regulation of AI technologies, but that this would require better public communication, education, and real empowerment to be effective.

> "Individual privacy actually has a huge chilling effect on sharing useful data, but it seems to have no effect – no detectable effect – on preventing these powerful organisations from accumulating more and more and more."
>
> (UK expert interviewee)

Discussion

When we look at privacy, agency and trust through an intercultural lens, we find very different underlying ways of seeing people, actions, institutions, and the world both between societies and even within the same social and cultural context. At the same time, we can also perceive convergences, for better or worse. These are often driven by shared neoliberal and/or democratic worldviews, similar political-economic interests, systems and institutions, global flows of capital, information, labour and technologies, and the imperative for interoperable legal and regulatory frameworks and technological standards to facilitate international trade.

In this context, the task of constructing systems of global ethical principles and governance for AI is a complicated affair: we come back to the question of whose values are represented, how these are interpreted across linguistic horizons and dissimilar forms of life, and how and to what extent different views of the world can be reconciled, or even brought into common understandings. Striking the right reconciliation and including the diverse perspectives of a globally representative set of stakeholders—especially those who have been marginalised or rendered vulnerable by digital transformation and datafication—are vital preconditions for achieving just outcomes. Such a dialogical, conciliatory, and multi-stakeholder approach would ensure that Anglo-European and American values are not simply imposed on a diversity of other societies via hegemonic policies, standards, governance regimes, and technological means in a new form of "informational colonialism".

The PATH-AI interviews and surveys conducted to date have focused on case studies from health and care, but many of the findings cut across other sectors. One aspect that unites both the UK and Japanese findings is how privacy, agency, and trust are bound together in the context of emerging AI and data-intensive digital technologies by growing asymmetries – of data, information infrastructures, informed choice, resources, capabilities, and ultimately power – between users (whether citizens, data subjects, or consumers), governments, and companies. These asymmetries often left respondents feeling confused and disempowered. There was a pervasive sense of opacity and a feeling of distrust among members of the general public in both countries when it came to the growing number of black-box mobile apps, online tools, and algorithmic systems where it was not clear to users what these technologies were doing, how or even whether they were working, what the future implications of data being extracted might be, and so on.

The recent emphasis on individual privacy rights and data protection, and the industry that has grown up around them, could be interpreted in part as one attempt to push back and level the playing field in some way, with users asserting agency to avoid being seen in the same way that governments – to some extent – and companies – to a much greater extent – obscure their inner workings from citizens, data subjects, and consumers. The trend to privacy also seemed to reflect profound and growing societal distrust in governments and technology companies alike.

However, many experts argued that the implementation of data protection legislation to date had tended only to confuse citizens while not preventing companies from collecting ever-greater amounts of personal data. Indeed, there was a sense from respondents in both the UK and Japan that citizens, government, and regulatory and legal systems were simply not able to keep up with the rapidly increasing complexity of technological developments, leading to even larger and growing knowledge and power imbalances while creating widening zones of regulatory ambiguity and under-capacity – both of which favoured technology companies.

The way in which governments and technology companies make top-down decisions on behalf of citizens, data subjects, or consumers, often without fully informing them in a way that they can clearly understand, while centrally controlling digital tools and their databases, can perhaps best be described as "technocratic paternalism". This strategy was viewed by many respondents as rendering users disenfranchised and passive, with little sense of participatory parity or agency. Several argued that this mode of decision-making reflected governments and tech corporations acting unilaterally in a situation of widespread public distrust. At the same time, it served to further fuel this distrust, as well as suspicions about hidden motives and a sense of disempowerment among citizens and data subjects. Across both Japan and the UK, experts instead argued for the need to embrace a collective and more horizontal approach to these societal-level issues that went beyond governments acting in a top-down, paternalistic way to make increasingly complex and consequential decisions on behalf of citizens and data subjects. Instead, they call for public consultations, communications, meaningful participation and, in the Japanese context, "social consensus" on the way forward as societies, taking into account collective processes of negotiating costs and benefits, rather than simply devolving decision-making to individuals.

In order to address these power asymmetries, and to build back trust in government, companies, and the technologies they are being presented with, many respondents said they wanted far clearer explanations of and public discussions about what these increasingly complex technologies are, what they do, when they are deployed, how data is being collected and used, and how governance is taking place or should take place. Given significant disparities in digital literacy, many also called for far better public education. Research participants generally saw public involvement in these conversations as crucial to empowering them to contribute to, and play a part in deciding, the future direction of rules governing these technologies. Many respondents also called for more agency over what happened to their data, and for more tools to enable them to exercise this control. In this sense, governance of AI and other data-driven technologies, as well as a more active role in the design and development of new technologies, was understood by many research participants as offering the opportunity to begin to rebalance configurations of power. At the same time, many experts suggested that regulation could only go so far to governing these technologies, and that a variety of approaches was necessary in addition to significantly boosting regulatory capacity.

While many of these views had widespread agreement among research participants in both the UK and Japan, there were also important differences. On one level, the significant difference in health care systems led to differing concerns: in the UK, many interviewees worried about the potential for the ongoing privatization of the NHS by the back door, with tech companies outmanoeuvring and outcompeting it through their larger-scale collection and synthetic analysis of all kinds of data. In Japan, the greater concern was that the country was falling behind less scrupulous global competitors due to siloed and inefficient bureaucratic structures and an overly cautious attitude from the private sector. At a more fundamental level, a further notable difference was the divide in how AI and related technologies were viewed, with Japanese respondents more optimistic and seemingly more comfortable with the idea of, for example, interacting with robots – even though, perhaps surprisingly, they were less likely to have downloaded a DCT app, or to have used a symptom checking tool, or to have heard of or interacted with care robots.[7] In broad terms, while UK participants worried that emerging technologies might cause harm due to inadequate design and implementation, many Japanese participants seemed more worried that they would work *too well* or become too powerful, introducing dangers of lack of control or the creation of a future society ruled by what one respondent described as "extreme rationalism": systems of automated decisions and action that would be less easily contested.

By looking at these case studies in the UK and Japan through an intercultural lens, PATH-AI provides a starting point for reflection on how the situatedness of privacy, agency and trust in complex webs of social, political, economic and technoscientific relationships may help us to identify the salience of converging and differing values between cultural contexts. PATH-AI broaches considerations of what can be learned by taking such an integrative approach, and how and to what extent such confluent or divergent values could be reconciled in a more global view of AI ethics and governance. In the next stage of the project, we aim to develop and pilot a methodology for the intercultural co-design of a framework for more ethical and equitable human-AI ecosystems.

7 In the latter two cases, this may in part have reflected the difference in methodology, with Japanese participants completing written questionnaires while the interviews with UK participants afforded slightly more flexibility in asking and answering these questions.

References

Ada Lovelace Institute. 2020. “No green lights, no red lines: Public perspectives on COVID-19 technologies.” https://www.adalovelaceinstitute.org/report/covid-19-no-green-lights-no-red-lines.

Adams, Andrew A., Murata, Kiyoshi, and Yohko Orito. 2009. “The Japanese sense of information privacy.” *AI & Society* 24 (4): 327-341.

Allison, Anne. 2014. *Precarious Japan*. Duke University Press.

Angwin, Julia, Larson, Jeff, Mattu, Surya, and Lauren Kirchner. 2016. “Machine bias.” *ProPublica*, May 23, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Bannerman, Sara. 2018. “Relational privacy and the networked governance of the self.” *Information, Communication & Society*. https://doi.org/10.1080/1369118X.2018.1478982.

Bauer, Paul C., and Markus Freitag. 2018. “Measuring trust”. In *The Oxford Handbook of Social and Political Trust*, edited by Eric M. Uslaner, 15-36. Oxford University Press.

Bennett, Colin, and Charles Raab. 2006. *The Governance of Privacy: Policy Instruments in Global Perspective*. MIT Press.

Bensinger, Greg. 2021. “Americans Actually Want Privacy. Shocking.” *New York Times*, May 20, 2021. https://www.nytimes.com/2021/05/20/opinion/apple-facebook-ios-privacy.html.

Bridges vs SWP (South Wales Police). 2020. R (on the application of Edward Bridges) v. The Chief Constable of South Wales Police and the Secretary of State for the Home Department. Court of Appeal, Civil Division, case C1/2019/2670. https://www.judiciary.uk/wp-content/uploads/2020/08/R-Bridges-v-CC-South-Wales-ors-Judgment.pdf.

Cadwalladr, Carole, and Emma Graham-Harrison. 2018. “Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach.” *Guardian*, March 17, 2018. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

Crawford, Kate. 2021. *The Atlas of AI*. Yale University Press.

Doi, Takeo. 1973. *The Anatomy of Dependence*. Kodansha.

Edelman. 2021. “Edelman Trust Barometer 2021.” https://www.edelman.com/trust/2021-trust-barometer.

Erikson, Erik H. 1959. “Growth and crisis of the healthy personality.” In *Psychological Issues: Selected Papers*, edited by Erik H. Erikson, 51-107. International Universities Press.

Eubanks, Virginia. 2018. Automating Inequality: *How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin’s Press.

Frumer, Yulia. 2018. “Cognition and emotions in Japanese humanoid robotics.” *History and Technology* 34 (2): 157-183.

Fukuyama, Francis. 1996. *Trust: Human Nature and the Reconstitution of Social Order*. Simon and Schuster.

Gal, Danit. 2020. “Perspectives and approaches in AI Ethics.” In *The Oxford Handbook of Ethics of AI,* edited by Markus D. Dubber, Frank Pasquale, and Sunit Das. Oxford Handbooks.

Gygi, Fabio. 2018. “Robot companions: The animation of technology and the technology of animation in Japan.” In *Rethinking Relations and Animism: Personhood and Materiality*, edited by Miguel Astor-Aguilera and Graham Harvey, 94-111. Routledge.

Hao, Karen. 2021. “How Facebook got addicted to spreading misinformation.” *MIT Technology Review*, March 11, 2021. https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/.

Hern, Alex. 2021. “WhatsApp loses millions of users after terms update.” *Guardian*, January 24, 2021. https://www.theguardian.com/technology/2021/jan/24/whatsapp-loses-millions-of-users-after-terms-update.

Hildebrandt, Mireille. 2015. *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*. Edward Elgar Publishing.

ICO (Information Commissioner’s Office). 2017. “Royal Free – Google DeepMind trial failed to comply with data protection law.” https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law/.

ICO (Information Commissioner’s Office). 2019. “Information Commissioner’s Opinion: The use of live facial recognition technology by law enforcement in public places.”https://ico.org.uk/media/about-the-ico/documents/2616184/live-frt-law-enforcement-opinion-20191031.pdf.

IEEE (Institute of Electrical and Electronics Engineers). 2019. *Ethically Aligned Design (First Edition)*. https://ethicsinaction.ieee.org/#ead1e.

Kuldova Tereza. 2020. “Imposter Paranoia in the Age of Intelligent Surveillance: Policing Outlaws, Borders and Undercover Agents.” *Journal of Extreme Anthropology* 4 (1): 45-73.

Lukács, Gabriella. 2010. *Scripted Affects, Branded Selves*. Duke University Press.

Lukner, Kerstin, and Alexandra Sakaki. 2017. "Japan's political trust deficit." *Japan Forum* 29 (1): 1-18.

Mansab, Fatma, Bhatti, Sohail, and Daniel Goyal. 2021. "Reliability of COVID-19 symptom checkers as national triage tools: an international case comparison study." *BMJ Health Care Inform* 28: e100448.

Nissenbaum, Helen. 2010. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.

Power, Michael. 1997. *The Audit Society: Rituals of Verification*. Oxford University Press.

Public Health England. 2020. "Disparities in the risk and outcomes of COVID-19." https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/908434/Disparities_in_the_risk_and_outcomes_of_COVID_August_2020_update.pdf.

Raab, Charles. 2012. "Privacy, Social Values and the Public Interest." In *Politik und die Regulierung von Information: Politische Vierteljahresschrift*, Sonderheft 46, edited by Andreas Busch and Jeanette Hoffman, 129-151. Nomos Verlagsgesellschaft.

Rotter, Julian B. 1967. "A new scale for the measurement of interpersonal trust." J*ournal of Personality* 35 (4): 651-665.

UK Government. 2021. "UK launches data reform to boost innovation, economic growth and protect the public." September 9, 2021. https://www.gov.uk/government/news/uk-launches-data-reform-to-boost-innovation-economic-growth-and-protect-the-public.

UNESCO. 2021. "Recommendation on the ethics of artificial intelligence." https://en.unesco.org/artificial-intelligence/ethics.

University of Oxford. 2020. "Digital contact tracing can slow or even stop coronavirus transmission and ease us out of lockdown." April 16, 2020. https://www.research.ox.ac.uk/article/2020-04-16-digital-contact-tracing-can-slow-or-even-stop-coronavirus-transmission-and-ease-us-out-of-lockdown.

Uslaner, Eric M. 2002. *The Moral Foundations of Trust*. Cambridge University Press.

Warren, Samuel, and Louis Brandeis. 1890. "The right to privacy." *Harvard Law Review* 4 (5): 193-220.

Westin, Alan F. 1967. *Privacy and freedom*. Atheneum.

Wright, James. 2021. “Overcoming political distrust: the role of ‘self-restraint’ in Japan’s public health response to COVID-19.” *Japan Forum*. DOI: 10.1080/09555803.2021.1986565.

Yamagishi, Toshio. 2011. *Trust: The Evolutionary Game of Mind and Society*. Springer Science + Business Media.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* Public Affairs.