



## Valuing lived experience and co-design solutions to counter racial inequality in data and algorithmic systems in UK's digital services

Aunam Quayoum & Mark Wong

To cite this article: Aunam Quayoum & Mark Wong (16 Apr 2024): Valuing lived experience and co-design solutions to counter racial inequality in data and algorithmic systems in UK's digital services, Information, Communication & Society, DOI: [10.1080/1369118X.2024.2331781](https://doi.org/10.1080/1369118X.2024.2331781)

To link to this article: <https://doi.org/10.1080/1369118X.2024.2331781>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 16 Apr 2024.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

# Valuing lived experience and co-design solutions to counter racial inequality in data and algorithmic systems in UK's digital services

Aunam Quyoum and Mark Wong

Urban Studies, University of Glasgow, Glasgow, UK

## ABSTRACT

In the United Kingdom, the COVID-19 pandemic has accelerated the movement towards large-scale digitisation of *everyday* services: from healthcare, social housing, social security to utilities. However, what is lesser known are the vulnerabilities minoritised ethnic citizens face in datafication processes and how they are racialised within data and algorithmic systems of these services. Its potential to amplify systemic racism and inequalities is concerning yet under-studied, particularly in the UK context. This paper argues it is imperative to understand, and value, racialised minorities' lived experience to inform and improve digital services' design to be more inclusive and equitable. Drawing on qualitative interviews and workshops with individuals who identify as a minoritised ethnic individual across England and Scotland, we explore people's lived experience of everyday digital services. The findings highlight issues related to trust, data privacy, and poorer quality access to services. Such experiences are shaped by the fears and lived experience of racism, both structural and institutional. This paper concludes the use of co-design methods can strengthen capacity amongst racialised communities and stakeholders to articulate where inequities are occurring, understand how to counter harm, and co-create solutions. We outline our case for a co-design approach to guide public and private sectors' decision-making and policies, so digital services are equitable and responsible by design.

## ARTICLE HISTORY

Received 14 December 2023  
Accepted 8 March 2024

## KEYWORDS

Co-design; racial justice; digital services; lived experience; anti-racism

## Introduction

While we transition into an increasingly digital present, the lived experience, and concerns raised by people who are adversely racialised and affected by systemic racism and white privilege in western societies, such as the United Kingdom (UK), are most often ignored and under-valued in the development of artificial intelligence (AI) and digitisation. Inclusion, representation, and diverse voices in the AI and data transformation space is thereby pressingly required (Benjamin, 2019). There is urgent need for a

**CONTACT** Mark Wong  Mark.Wong@glasgow.ac.uk  University of Glasgow, 25 Bute Gardens, Glasgow G12 8RS, UK

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

better understanding of the mechanisms and processes we can rely on, to make better decisions on why, how, and when, AI and digital systems should be, or should not be, used and whether they can serve different communities in equitable ways. This paper seeks to challenge the privilege of being able to ignore the racial bias and discrimination baked into everyday digital services, by centering the lived experience of people most negatively impacted by the unfairness and injustice embedded, and reinforced, through data and algorithms.

The harms and bias embedded in data and AI systems used in digital services has been widely documented in a myriad of disciplines, including media and communication, science and technology studies, social policy, education, computer science, philosophy, and AI ethics (Crawford, 2021; Eubanks, 2018; Li, 2021; Noble, 2018; O’Neil, 2016; Williamson et al., 2023). This global debate provided much-needed evidence to call for change in the public and policy discourse to strengthen regulation on AI internationally, such as the United Nations high-level AI advisory group and the European Union AI Act (European Commission, 2021; Fung & Etienne, 2023). However, despite public appeal for greater fairness, transparency, and equity, there is limited consensus on how this can be achieved (Duke, 2023), while the pace of change in policy and innovation systems remains slow, while data driven discrimination is rife.

Ample evidence in critical AI studies has revealed AI harms on racialised people, including biased facial recognition and racial profiling in predictive policing (Buolamwini, 2023). Racist algorithmic decisions in judicial systems and racial-targeting in immigration and borders also perpetuate prejudices. AI models, hence, cause harm by transmitting discrimination, toxicity, misinformation, and negative stereotypes (Kidd & Birhane, 2023). AI innovation ultimately entrenches a hierarchical system of discrimination and injustice deeply rooted in histories of colonialism (Adib-Moghaddam, 2023). However, what is lesser known is how people makes sense of and navigate these systems and harms, particularly in the UK context. Our evidence reveals nuanced realities of emotions, frustration, and hopes that racialised people have towards making digital services fairer and trustworthy. This paper, therefore, offers new understandings of lived experiences of racialised people’s interactions with digital services underpinned by unfair designs of data, algorithms, and AI, filling an urgent gap concerning what change people hope for, and why. By highlighting co-design as a desired path towards realising change and justice, this paper offers a leading contribution to growing debates on the importance of centering the role marginalised communities play in data and AI and amplify the voice of those most impacted (D’Ignazio & Klein, 2020). Our contribution thereby adds a new perspective to what Birhane (2021) theorises as a ‘relational ethics’ approach; addressing algorithmic injustice by and for those who are disproportionately impacted.

A co-design approach of relational ethics in research offers space for the voice and expertise (by learning and/or by experience) of minoritised ethnic people to be centred and visible in the design of these systems, ensuring they do not harm and discriminate against minoritised ethnic people. Co-design has a significant potential, when used carefully and equitably, to ensure the priorities and design of these services are co-created and co-determined by the people most impacted or at risk of bias and poorer health and social outcomes (Ada Lovelace Institute, 2021a). Our recommendation for co-design approaches builds on qualitative evidence collected through interviews and workshops with minoritised ethnic people across England and Scotland. We suggest the concerns

and issues highlighted from this evidence underlines the need for service providers and digital system designers to use co-design, to ensure digital services are equitable, trustworthy, and fair.

Although this paper primarily addresses the UK context, it also offers much-needed evidence to inform change in policy and practice in a space where political and public interest is increasing across the world about digital futures and AI ethics. We also want avenues for international research and policy to be encouraged through this paper, as it builds on an important intellectual agenda for many scholars, especially those who are often silenced by so-called ‘frontier AI’ innovators in western countries who assume critiques to be hinderance instead of *progress* towards better innovation in just and equitable ways (Buolamwini & Gebru, 2018).

### ***Digitisation of everyday services***

In the UK, the digital transformation of key services during the COVID-19 pandemic rapidly accelerated. Access to health services, for example, including appointment booking, consultation with General Practitioners, ordering repeat prescriptions, the COVID-19 ‘track and trace’ app, the vaccine certification badge, are a few stark examples of how rapidly everyday services shifted from being primarily in-person to being digitalised. This is also seen elsewhere including social housing (housing applications), energy services (customer service, bill payments, and smart meters), and welfare benefits systems (Currie & Podoletz, 2023). Although this is not a new phenomenon, the COVID-19 pandemic forced service providers to ramp up the pace of digitisation and, at times, remove non-digital support entirely. But the consequences of this expedited digitisation sits alongside the structural racism and inequalities services that are embedded, combined with processes of racialisation through datafication, are concerning but the impact of is lesser known in the UK context (Wong, 2022).

Since then, increased number of algorithmic and automated decision-making systems have been deployed across a wide variety of domains of everyday services, including deciding who gets welfare benefits, triage of access to care, pilots in health diagnosis and image scanning using computer vision (Ada Lovelace Institute, 2022). There are also other applications of AI across the world, such as border control and visa applications (Dumbrava, 2021), screening of job applications (Crawford, 2021), and behavioural modelling in insurance and loan premiums, which evidence shows perpetuates racial profiling (Favaretto et al., 2019). This raises a problematic path of digitisation of services and datafication of people. We currently have little understanding of the experiences of minoritised ethnic people when they interact with these data and algorithmic systems in the UK context and what we can do to mitigate these impacts.

### ***Racialised data systems***

Algorithmic tools have led to certain minoritised ethnic communities to be wrongfully denied services such as housing, access to credit, targeted as ‘sham’ marriages, rejected from social benefit claims, labelled as clinically vulnerable to disease, as well as over-policing and surveillance because of data banks such as the London Metropolitan police’s

unlawful ‘Gang Matrix’ (Amnesty International, 2018; Couchman, 2019; O’Brien et al., 2022). Against this backdrop, minoritised ethnic individuals also report low levels of trust in the technical competence of services, including the NHS with concerns around data handling, and fears around the sharing or selling of personal information to private companies (Wilson et al., 2022). This is further amplified for racialised communities by a hostile environment agenda, which has seen British Black-Caribbean citizens wrongfully facing deportation based on administrative errors by the Home Office (Taylor, 2022), and is also enacted via the NHS.

As such data and AI systems are amplifying systemic racism. Yet there is limited research to understand the day-to-day vulnerabilities that minoritised ethnic people face, particularly where access to these services are critical to livelihoods and wellbeing. There is limited understanding regarding the effects of facing these everyday digital experiences, which are often difficult to make sense of – even more so given the opacity of the underpinning algorithmic systems. Too often, it is unclear to people what is happening behind the screen and why they are not able to access the services and outcomes they expect to, when interacting with services digitally. This paper seeks to address this gap and highlights the frustrations and reflections that minoritised ethnic people have about digital services, specifically in health and social housing.

## Methodology

This paper adopts a qualitative methodology, drawing on a thematic analysis of semi-structured interviews and two co-design workshops, conducted as part of a research project focused on the digital experiences of minoritised ethnic communities. During May 2022–March 2023, 100 semi-structured interviews were conducted. The interviews include individuals who identified as a minoritised ethnic person in four locations in England and Scotland, where a high proportion of minoritised ethnic people live. Participants were interviewed if they self-identified with one of the ethnicities targeted. These are the largest minoritised ethnic populations in the UK, including: Black African, Indian, Pakistani, Black Caribbean, Bangladeshi, and Chinese. We partnered with several community organisations to recruit participants locally with the aim of targeting individuals from a range of ages, minoritised ethnicities, levels of digital access and/or use of services such as social housing. Five community organisations helped with advertising the study to their members through internal communication channels, using recruitment posters provided by the project. The recruitment materials were translated into four languages as requested by organisations (Urdu, Gujarati, Bengali, Simplified Chinese).

The data presented here is based on a coalescence of larger themes identified from the 100 interviews, but through the specific narratives of 6 selected individuals. The 6 interviewees selected for discussion had some independent engagement with digital services. They present as case study ‘spotlights’ which illustrate the three inextricably linked themes related to the perceptions of minoritised ethnic individual responses to digital services, which include; (i) the experience of race and racism, (ii) data privacy, and (iii) feelings of mistrust towards institutions and service providers (Yin, 2008).

Below is a breakdown of the interviewees discussed in this paper referring to residential location, age, gender and ethnicity as defined by the participants themselves.

- Participant 1: Glasgow, 41, Male, British Pakistani
- Participant 2: Glasgow, 40, Male, Scottish Pakistani
- Participant 3: Tower Hamlets (London), 38, Female, Pakistani
- Participant 4: Tower Hamlets (London), 38, Female, British Bangladeshi
- Participant 5: Manchester, 42, Female, Black African Caribbean or Black Other<sup>1</sup>
- Participant 6: Manchester, 53, Female, Black British

We acknowledge the participants selected for this paper fall within the middle-aged bracket, and therefore the interpretation of the data is skewed towards the lived experience of middle-aged minoritised ethnic people. The middle age group was defined by the project as individuals between the ages of 35–60. Whilst this could be seen as a limitation, it may also be indicative of the nature of accruing lived experience of racism and structural inequality over time, and as more engagement with services is needed in this age bracket compared to younger generations. Further, it may be that middle-aged individuals are more likely to assume or be handed the responsibility for navigating digital services, with all participants bar participant 5, bearing additional responsibilities to support others (families and relatives) to access digital services, such as booking GP appointments, applying for social housing and managing online energy accounts.

Interviews were conducted both online via Zoom and in-person on a one-to-one basis depending on participant's preference. Interviews were conducted by members of the research team, all of whom also identify as minoritised ethnic individuals. Participants were asked specific questions related to their experiences of digital health, social housing and energy services, but were also asked questions about perceptions and experiences of online harm, data privacy and issues of racism and discrimination where relevant.

Interviews were audio recorded (where permission was granted by participants) then transcribed verbatim by a transcription service. The data was de-identified and participants were assigned a unique code. The data was analysed using a thematic analysis approach; iteratively coding relevant themes using NVivo 14 and immersion in the data. The study received ethics committee approval from Heriot-Watt University, as lead institution of the project.

In addition to interview data, this paper draws on data collected from two co-design workshops, conducted in October 2023, to supplement the themes identified from interviews and allow us to consider what next in how to mitigate issues of inequity. Workshops were held in Birmingham and Glasgow, designed in partnership with local community organisations. Workshop participants included stakeholders and representatives from across health, third sector, housing, minoritised ethnic community members as experts by experience, design consultancy, public sector, and local authorities. A total of 17 participants in Glasgow and 13 in Birmingham contributed to the workshops (Table 1).

**Table 1.** Workshop participants.

	Birmingham	Glasgow
Community	10	4
Stakeholders*	3	13

\*Stakeholders participated include a majority of people also with lived experience of racialisation.

Community participants were only asked to represent their own views and lived experience, so their participation was not intended as a form of representative sampling nor to infer what particular communities think. It is their individual lived experience as a racialised person which is the focus of our analysis of the interviews and workshops. The analysis presented in this paper is, therefore, supplemented by data arising from the workshops' dialogues and deliberations. The data was analysed based on their relevance to the three themes identified from the interviews and, additionally, the need for the co-design of digital services. What we have heard at these workshops supports the case that more of this type of deliberative research and discussion is valued and needed in the digital space. Each workshop lasted 3 hours, providing rich data to complement the interview analysis. In the Glasgow workshop, there was continuity in representation of community members by selecting those who had participated in our interviews. Therefore, although the number of interviews included in this paper might seem small, by combining different sources of data, this paper offers rich insights and in-depth narratives as the basis of our rigorous analysis. Ethical approval to conduct the co-design workshops was granted by the University of Glasgow.

## Results

Out of 100 interviews with minoritised ethnic individuals, 49 participants used, or were supported to use at least one digital service for health, social housing and energy. The rest either did not engage with any digital service because face-to-face options were still available, responsibilities were delegated to others, or they refused to change how they have historically accessed a service. Interview data from minoritised ethnic individuals illustrate a range of responses towards the digitisation of everyday services. Responses varied from: the acceptance of a poorer service, or withdrawal from a service, resistance to engage digitally or provide data, and at times, feelings of coercion to comply to data systems. Such responses were influenced by a variety of inextricably linked experiences and feelings which are thematically presented in this section as: (i) the experience of race and racism, (ii) data privacy, and (iii) mistrust towards institutions and service providers. While perceptions may be shaped by intersecting factors related to socioeconomic status, age, gender, language and health needs, we have purposefully grounded data analysis through the lived experience of race/ethnicity. The reason for this is because the research focus is on the collective experience of racial minoritisation and how this is realised within seemingly 'colour blind' data systems, but which are known to embody systemic racism (Feagin, 2013).

### *The experience of race and racism*

This theme relates to the structural, institutional and inter-personal effects of racism on people who are minoritised because of their race/ethnicity/religion and/or citizenship status in the UK. Some participants reflected on structural inequalities in British society or experiences of racism, which shaped perspectives on digital services. During interviews these experiences were often sparked by questioning related to ethnicity information or experiences of feeling unheard by services. Participant 5 offers an example of the deeply rooted racial inequities that persist in health:

‘A recent report came out that said ethnic minorities are disproportionately receiving worse services from the NHS, waiting 61+ weeks and white counterparts 54 weeks – it’s institutional racism! ... How can we live in a society with that disparity ... maternity deaths for black women – no one knows why this is happening, they say ‘oh it could be this, it could be that’ but nobody actually knows!’ (Participant 5)

This individual reflected on the issue that while data evidences racial inequity in health services, rarely is such data supported with explanations of why, which ultimately hinders the eradication of disparities across services. This led to them trying to avoid digital services in health and sharing data online. Participant 5 explains:

‘It’s never clear why the ethnicity information is asked. Why?! Even on ... really long NHS surveys they ask you to fill in. They don’t really explain why! I will only tell them so much if I really, really have to. Likely to say Black other. [I] Want to know the context of the question. I’m very choosy. I’m aware of online applications for benefits being sabotaged, small elements in the DWP department [for work and pension] that were racist to prevent and delay applications amongst minority ethnic people. So I’m very aware of how information could be used against you.’ (Participant 5)

The continued experience of structural and institutional racism embedded in services, such as the NHS, is leading some minoritised ethnic individuals to question services and is being carried over to people’s view on engaging with digital services. Context matters and so the concerns minoritised ethnic people have about systemic and institutional racism, cannot be disentangled from the need to distrust services as whole, as a safety response, which likely includes digital services.

Reflecting on their social housing situation, participant 4 describes delays in their housing claim which saw them living in ‘limbo’ in temporary accommodation, with limited human contact to offer updates on the progress of housing applications. Despite the accommodation being a half-demolished, tower block, prone to break-ins and vandalism, it was a space they had to inhabit with their young family.

‘Once me and my daughter were coming into the building, and it just missed us, a big slab of concrete fell right in front of us, and I had to report that. I’d call them first, but they’re saying, you need to give email ... But there’s been numerous emails that I’ve sent because once I became homeless ... my account for bidding was suspended for two years ... then two years they haven’t sorted it out. I’ve tried to chase them. I’m sitting in a temporary accommodation. I don’t know what’s going to happen to me. I can’t even bid for any house ... You’re in the limbo with everything. At least face-to-face you get to speak to someone.’ (Participant 4)

Meanwhile, they were also responsible for managing multiple people’s access to digital health services because of barriers for others related to English language or digital skills. This was compounding their mental health negatively. The temporal unknown of the situation also meant they were reluctant to spend what money they had on the high cost of a rolling or fixed term internet contract, and instead had to rely on mobile data to access digital services. The participant felt invisible, speculating whether they were being racially othered by the social housing provider which could be easier to do via an online system.

‘My assumption at the moment is that ... am I being put in a back burner because of the background I am? ... Is that why I’m not being looked into? ... It’s almost like, ‘Oh, it’s



one of *you* again.’ It’s like they brush everyone with the same brush but they don’t look at individuals and their individual situations’ (Participant 4)

They expressed that despite living in an area with a larger population of Bangladeshi residents in Tower Hamlets, they felt their social housing service perhaps cared less about them as a result of being the ‘majority’ minority. This was felt to have impacted the sense of urgency and attention given to their housing situation.

The data indicates the perceptions minoritised ethnic people have about digital services are influenced by fears of racism or experiences of racism, personally and in historical institutional contexts. While none of these experiences are exclusive to digital life, it sets an important context which services and institutions must be mindful of when trying to develop newer or extend their digital services. The continued inaction to address issues of racism will likely cause significant concerns amongst minoritised ethnic people towards digitisation and datafication of services. Ultimately, people are concerned about being othered by digital systems and the adverse impacts of racialisation being deepened and perpetuated *through* these systems.

### **Data privacy**

Lived experience of racism and poor prior experiences with key services subsequently influenced attitudes towards data privacy and information security. Participant 2 described a general suspicion and reluctance to share information irrespective of which service or institution, based on their experience of racial profiling, principally with police services in both Scotland and England.

‘Guess how many times I’ve been stopped? 48 times! So I have to base it on what is it about me that’s making the police target me? Then obviously it must be either racism or it must be that I look different from them ... I’ve been asked to do all sorts of stuff on many an occasion that is a breach of my privacy. A personal breach of my privacy! So I’m reluctant to just give my name out willy-nilly or my details out willy-nilly.’ (Participant 2)

For them, ‘sensitive’ personal information stemmed beyond ethnicity, to include name, location and even statistics related to energy usage leading them to invest in data privacy tools to block, delete and do-not-track for online activities. They only participated in this research and reluctantly signed a consent form, based on the positive prior interaction with the interviewer, who was a person of colour.

Their ever conscious awareness for data privacy extended to algorithms and smart energy services with concerns around the ‘handing’ over of mass personal data for the perceived benefit solely of private companies.

‘Algorithms are created by individuals of a certain persuasion, not by – or they’re not fair, I’ll say that much. I know that for a fact, so I am concerned about algorithmic bias ... Much like with the police, I will not disclose anything online. I would much rather go and disclose information face-to-face where I can.’ (Participant 2)

More broadly across interviews, participants expressed mixed concerns regarding the sharing of ethnicity data and fears around how it may be used by services and institutions. While one recounted the absurdity of being asked for their ethnicity in a credit card application, others saw some benefit for services regarding equality monitoring,

while others felt that providing ethnicity or religion information could lead to negative stereotyping that could be delivered by an 'objective' digital system as opposed to a 'human' interaction.

Some reflected on the complexity of ethnicity labelling, feeling the need to caveat their responses with explanations of being born in the UK or having lived here for most of their lives. While reflecting on responses to ethnicity questions, participant 5 expressed concerns of bias and discrimination, which had some influence in how they answered questions.

'I avoid sharing ethnicity information when I can, not 100% confident about where my information will be used and shared and if it will be taken in a positive perspective. I just say that I'm Black African Caribbean or Black other so it's a bit purposefully vague.' (Participant 5)

In light of this, digital services must provide reassurances to racialised minorities in the UK about the safety and use of their data – which includes ethnicity, religion, citizenship status. The collection of these data types does not exist in social isolation from the wider bias, discrimination and hostile environment towards racialised minorities in the UK. Further still, we know that beyond ethnicity categorisation labels there are other data points which are known to produce biased decisions for racialised minorities via proxies such as postcodes or geographical data. Discrimination legislation can be difficult to account for proxy variable discrimination, meaning it is necessary to engage in discussions with people and services that is deliberative and tries to foresee and mitigate such potential harms from the outset of the design of data systems (Wachter & Mittelstadt, 2019).

### ***The absence of trust and mistrust***

As previously outlined, minoritised ethnic people have varying levels of trust towards digital services, which is inextricably linked to the lived experience of racialisation. Feelings of mistrust were often articulated through questions about why information is collected by services, how it will be used and who it will be shared with. Of the 100, 43 participants expressed general concerns, while 23 participants specifically reflected on providing ethnicity data. There was a quandary amongst some participants about whether providing ethnicity information will lead to biased targeting and decision making and may be more difficult to challenge in a digital space, or if the refusal to provide information may be seen as suspicious behaviour or hinder effective race equality monitoring. These thoughts sometimes sat alongside the knowledge of how information on personal characteristics may be used with good intent but still carried some discomfort:

'I have a family member who worked in the NHS, and she says they need that data to show that they're meeting quotas, but sometimes I just feel like it's just ... I get the purpose of it. Sometimes, I don't feel comfortable giving out ethnicity and those kinds of things. They say it's anonymous ... but it's almost that thing that if they know where you're from, it could easily sway your decision, despite all the reassurances.' (Participant 3)

Research on the views of minoritised ethnic people towards data in health and social care services shows general mistrust, which is amplified by people rarely seeing positive outcomes from data sharing or not kept informed about what happens with their data

(Wilson et al., 2022). Ethnically minoritised citizens in the UK can be reluctant to participate in data sharing because the potential benefit for them, their communities, or the wider public is unclear (Mac Manus, 2021). However, research has also shown that Black respondents are significantly more likely than any other ethnic group to feel comfortable in sharing personal data for the purposes of checking services are fair to individual groups (Centre for Data Ethics and Innovation, 2022). Therefore, greater clarity around the purpose of data collection needs to be communicated more accessibly, to build trustworthiness.

Research suggests trust in the NHS is higher compared to government, social media or private companies (*ibid*). However, prompted by discussion of NHS apps and how problematic the ‘track and trace’ system was during the COVID-19 response, one participant explained:

- Interviewer: ‘It doesn’t matter that if it’s the NHS or it’s Google, you don’t distinguish between who you might trust a bit more than the other?’
- Participant 6: ‘As far as I’m concerned they’re all joined, they’re all working together. I don’t want anyone tracking me or anywhere I go. Simple as ... Specifically I’m concerned that – everybody thinks I’m crazy when I say things like this, but first of all, the powers that be, the higher-ups, I think want to kill *us* [gestures towards interviewer] all. That’s what I think.’
- Interviewer: ‘So you feel the state is violent? ... Towards people of colour?’
- Participant 6: ‘Yes, and unfair. I’m going to go as far as to say *scared*, as well.’

They discuss the fear the ‘powers that be’ (the state) have towards ‘black consciousness, black power, and what *we’re* really capable of’ as they gesture towards the interviewer and themselves. The negative experience of one NHS digital service also had some influence on their feelings towards other digital services such as smart energy meters. The only digital service they forcibly had to use is the social benefits system as a full-time carer for a child with a disability.

Beyond digital health services and data collection, participants also described general mistrust in the motivations behind a digital service altogether whether a public or private actor, with perceptions of collusion. Examples of NHS partnerships with Google’s DeepMind or private data analytic companies such as Palantir to handle patient data (Hall, 2020; Hern, 2017), illustrates the limitations of tracking public trust by stakeholder or institution (CDEI, 2022). Rather there is a blurring of parties involved in digital services and as the state participates in these data systems, it is seen to legitimate opacity and limit accountability (Pasquale, 2015). This creates, if not can perpetuate suspicion and mistrust in the motivations behind the *purpose* of digital systems altogether and who ultimately benefits from this transition.

In light of concern digital services are capable of capturing more personal data that could facilitate discrimination, participants reflected on inconsistently responding to identity questions based on situated judgements related to perceptions of safety, trust and prior experiences. Many preferred broader ethnicity labels such as ‘Asian’ or ‘Other’ as opposed to Asian-Bangladeshi, for example. Some refused to provide information altogether. There were concerns that providing more granular ethnicity information could negatively impact access to services.

'It's like when I am filling out the application, I always class myself – I don't say, 'I'm Pakistani.' I say, 'I'm British/Pakistani' ... As soon as you put, 'Pakistani,' from there it's like, 'Right, okay, mate. Your chances are being brought down.' I've noticed that in Scotland ... especially with this housing thing.' (Participant 1)

The variability of responses to ethnicity questions can present challenges for effective monitoring of equality, fair access and outcomes, as services will subsequently have a partial understanding of service users. While we maintain individuals have the right not to share personal information as a privacy protective tool (Coopamootoo, 2020), this presents challenges for algorithms and machine learning models as 'pipeline' or training data is relying on skewed, partial or inaccurate datasets (Veale & Binns, 2017). Ethnicity as a social categorisation is politically and culturally situated and digital systems can undermine this complexity. As such, collecting and utilising ethnicity information can be a double-edged sword – no information means we are less able to hold services that may be discriminatory or biased to account, or audit algorithms for fairness (Costanza-Chock et al., 2022), yet an over reliance on ethnicity information can lead to the homogenisation of diverse people, and perpetuate racial stereotyping.

Discussing the red lines of data sharing with impacted communities is important given the uncertainty and lack of clarity individuals perceive and the extent to which institutions are held responsible. Yet solutions for trust are typically focused on how to improve accuracy or trust in the technology, instead of addressing wider questions of efficacy, power and legitimacy in society (Redden et al., 2020). For example, AI and trust often draw on the concept of explainability (Toreini et al., 2020) and communicating decisions and outcomes in an accessible way. However, it may not be practical, or relevant to explain all aspects of a digital system but rather focus on the *process* of decision-making (Ada Lovelace Institute, 2021a, p. 23). Thus, discussions on where, when and how digital and algorithmic systems should or should not be used, need to be discussed by a range of stakeholders within communities (particularly those most negatively impacted), with broadly agreed upon definitions of fairer decision-making and processes – which presents the case for co-design approaches.

## Discussion

The data presented in this paper evidences some concerns minoritised ethnic people have towards digital systems, the institutions they represent, the values they appear to embody and subsequent trustworthiness. Service providers must consider minoritised ethnic people's experiences of racialisation and racism when designing services, because although trust can withstand some testing, it is the cumulative impact of repeated failings and negative experiences that damage trustworthiness (Sheehan et al., 2021). The use of co-design methods in research and the design of services is increasingly important to consider in mitigating harms for negatively impacted groups, building trust and confidence in systems, but also as a method for auditing digital systems and algorithms (Costanza-Chock et al., 2022). We therefore outline our case for a co-design approach to guide public and private sectors' decision-making and policies, so digital services are equitable and responsible by design, rather than reacting to problems (Crawford & Schultz, 2014)

Co-design can be described as a form of community engagement, and broadly understood as an exchange of ideas, opinions and good practice (Wellcome Trust, 2011).

However, what distinguishes co-design approaches from general engagement activities is that invited persons are expected to play a larger role in shaping outcomes and developing solutions, rather than being passive recipients of information. Co-design models of engagement can present as a form of learning for all parties involved as knowledge is developed and deliberated on as a collective body, but with a commitment to producing some form of change (Bray et al., 2000). The Ada Lovelace Institute (2021b, p. 55) defines a co-design approach as a ‘collaborative model’ of governance which includes public, private and civil society representatives involved at various stages of the design and development of algorithmic systems. The benefit of co-design from a public sector, company or providers perspective is that it can facilitate greater space to communicate with the public, understand how they feel and begin to identify the terms of use, priorities and urgencies amongst different communities (Ada Lovelace Institute, 2021a).

During our project’s workshops, participants from minoritised ethnic communities and stakeholders both reflected on the value of inclusive research throughout the lifecycle of designing and testing digital systems – an opportunity which is often under-utilised or not at all.

‘What a lot of these services do, the problems come afterwards because they haven’t done this. It’s not a people approach. They haven’t used real people in the design and re-evolving it and iterating.’ (Birmingham workshop community participant)

The feeling here is that minoritised communities are not consulted, let alone collaborated with. The result of this are services which are problematic, inefficient for people and more costly for services to rectify. Feeling unseen or unheard by systems and institutions also has a negative effect on trust and will need concerted efforts by services to regain, particularly in a context where the need to distrust, amongst minoritised ethnic communities, is a method of staying safe from systemic racism (Evangelist, 2022). Participants from minoritised ethnic communities argued their lived experience and knowledge would be of great benefit to the design of digital services, which research has also suggested is *equally important* as technical knowledge and expertise (Ada Lovelace Institute, 2021b). Stakeholders too reflected on the value of co-design as a feedback loop, which allows services to be responsive.

‘Ensuring that there’s co-design at the beginning, at the earliest possible point with service users, and within that including a better feedback loop so that if a [digital] form or a service isn’t working for you, and it’s really annoying, that you’ve got somewhere to go to, and that those issues can be resolved.’ (Glasgow workshop stakeholder participant)

Designing policy and practice solutions through a co-design approach can signal an institution or organisations commitment to change and service improvement, which can contribute towards building trust (Data Justice Lab, 2021). Moreover, it can generate a feeling of ownership amongst communities and act as positive experience in itself (Bradwell & Marr, 2008). This is crucial to consider as this paper has shown the mistrust minoritised ethnic individuals have towards everyday digital systems. By providing steps to adequately inform and empower people with the knowledge of how digital systems operate, and how their data is captured and used, can lead some way in rectifying historical issues related to a lack of trust in public and private services.

However, co-design does require careful planning and negotiation of power relations to ensure who participates is inclusive, appropriate, and dialogue is deliberative, and solution-minded (Fung, 2006). Co-design also needs to be approached in a non-tokenistic way, or else it only perpetuates the exploitation and unequal treatment of racialised communities, by exhausting their labour, time, effort – and ultimately trust. Stakeholder participants reflected on engagement tokenism but also on the unfamiliarity of how to approach co-design meaningfully:

‘We need to ensure that the mechanisms and frameworks for engaging diverse voices in the development of digital services is easily available to service providers. A lot of them, they can have excuse of, ‘I don’t know how to do it.’ If research projects like this can come up with very clear guides, how to best practice in this. Lots of companies can also do it in a tick-box manner where it’s not engagement. It’s how do you get meaningful engagement in co-design in ensuring minoritised communities are represented in the rollout of any service.’ (Glasgow workshop stakeholder participant)

In our case, we worked closely with two community-based organisations in the cities where workshops took place. We made provision for adequate resourcing of staff time in those organisations to facilitate this work and input into the design of workshops. This facilitated knowledge exchange but also aided the recruitment of participants from local communities, representatives from services, voluntary and community-based organisations, and associated partners who align to the interests of the research and displayed a willingness to adopt changes.

We note there are challenges with co-design approaches, particularly when seeking to address issues experienced amongst ethnically minoritised groups. They are not homogenous groups and have a variety of experiences and priorities. Any community, therefore, can be better understood as ‘a group of individuals with power dynamics existing everywhere’ and through this awareness, we can begin to develop co-design activities that try to mitigate internal power dynamics and avoid over-simplifying a diversity of experience (Wellcome Trust, 2011). While there can be some degree of tension in the co-design process based on the divergent lived experiences we need to make sense of, the collaborative aspect is down to the conclusions and solutions arrived at together (Bray et al., 2000). Furthermore, while ethnicity is a complex identity marker for many, we can nonetheless argue that ethnically minoritised groups do share commonalities in experiencing marginalisation and racism and digital platforms and services do not exist in social isolation from those inequities.

## Conclusion

Digital systems are not socially neutral and risk reproducing inequalities for minoritised ethnic people. In this paper, we presented empirical data, offering insight into the harms and lived experiences of minoritised ethnic people in the UK, amidst the steep acceleration of datafication and digital services in everyday life. Underpinning these perceptions of harm are the deep rooted effects and fears of structural racism and inequity, which risk being replicated and amplified by digitising services. We articulate some risks that urgently need addressing before key services are digitised fully, and we use this analysis as a foundation to argue for the valuing of lived experience and co-design, which has

significant but under-utilised potential to drive improvements in digital service design and policies (McIntosh & Wright, 2019). This paper's evidence contributes to recent critical AI debates on the conceptualisation of how to centre the role of impacted communities in AI and data, building on Birhane's (2021) theory of 'relational ethics', to counter algorithmic injustice. We argue valuing lived experience through co-design with racialised communities is imperative to enact change. Our evidence also adds new insights into how the harms conceptualised in critical AI studies, such as discrimination, toxicity, and negative stereotypes, are perceived as being ever-present in everyday digital services. This paper, therefore, makes a leading contribution to concerns in global AI policies, such as the United States' National Institute of Standards and Technology, EU AI Act, and the G7 Hiroshima agreement, where a paradigmatic shift is emerging towards change being rooted within impacted communities. This article adds crucial evidence to this growing discourse, alongside seminar work such as D'Ignazio and Klein (2020) and Nelson (2024), to inform implementation of these policy principles.

We maintain that the use of co-design methods can strengthen capacity amongst communities and stakeholders to articulate where inequities are occurring, understand how to counter harm and to co-create solutions. Co-designing policy solutions and services can be challenging to organise and there are risks to how representative such approaches can be for diverse groups. However, when done with careful attention to issues of power and representation, co-design approaches can build trust through the process itself, making it worthwhile to do for groups in society most negatively impacted through datafication and who report lower levels of trust in digital systems. We acknowledge potential barriers of co-design due to the UK government's reliance on contracting-out and purchasing off-the-shelf systems from private companies, often as a cost-saving measure, but we believe the evident need for co-design demanded by racialised communities outweigh maintaining the status-quo of current practice. This paper offers an imperative and hope to change the ways that have always been done, but need not be. There is no quick fix in building trust amongst racialised minorities in the UK. However, creating spaces for deliberation through a co-design process can shift the predominant lens in which individuals are seen as passive consumers or end-users, to active contributors. This should be viewed as one positive step towards creating a more equitable digital landscape.

## Note

1. Participant 5 declined to be audio recorded and detailed notes were taken in lieu of by the researcher.

## Acknowledgements

The authors thank the project participants, the project team members, the special issue editors, and the participants of the York Algorithmic Symposium where this work was originally presented.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the UKRI Strategic Priorities Fund under Grants EP/W03235X/1, EP/W032333/1, EP/W032341/1, EP/W032058/1, EP/W032082/1.

## Notes on contributors

**Mark Wong** is a Senior Lecturer / Associate Professor in Public Policy and Research Methods and Deputy Head of Urban Studies at the University of Glasgow. His expertise is in racial justice and fair and responsible artificial intelligence and data systems.

**Aunam Quyoum** is a Research Associate at the University of Glasgow. She holds a PhD in Education and her research experience and interests align to themes of social justice and participatory methodologies.

## References

- Ada Lovelace Institute. (2021a). *Participatory data stewardship: A framework for involving people in the use of data*. <https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/>.
- Ada Lovelace Institute. (2021b). *Algorithmic accountability for the public sector. AI Now Institute and Open Government Partnership*. <https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/>.
- Ada Lovelace Institute. (2022). *Algorithmic impact assessment: a case study in healthcare*. <https://www.adalovelaceinstitute.org/wp-content/uploads/2022/02/Algorithmic-impact-assessment-a-case-study-in-healthcare.pdf>.
- Adib-Moghaddam, A. (2023). *Is artificial intelligence racist?* Bloomsbury.
- Amnesty International. (2018). *Trapped in the matrix: Secrecy, stigma, and bias in the Met's gangs database. Amnesty International United Kingdom Section*. [https://www.amnesty.org.uk/files/2018-05/Trapped%20in%20the%20Matrix%20Amnesty%20report.pdf?HSxuOpdpZW\\_8neOqHt\\_Kxu1DKk\\_gHtSL](https://www.amnesty.org.uk/files/2018-05/Trapped%20in%20the%20Matrix%20Amnesty%20report.pdf?HSxuOpdpZW_8neOqHt_Kxu1DKk_gHtSL).
- Benjamin, R. (2019). *Race after technology*. Polity.
- Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, 2(2), 100205. <https://doi.org/10.1016/j.patter.2021.100205>
- Bradwell, P., & Marr, S. (2008). *Making the most of collaboration: An international survey of public service co-design*. Demos in association with PwC's Public Sector Research Centre.
- Bray, J. N., Lee, J., Smith, L. L., & Yorks, L. (2000). *Collaborative inquiry in practice*. Sage.
- Buolamwini, J. (2023). *Unmasking AI*. Random House.
- Buolamwini, J., & Gebru, T. (2018). Transparency gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1–15.
- Centre for Data Ethics and Innovation. (2022). *Public attitudes to data and artificial intelligence tracker survey: Wave 1*. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1092140/Public\\_Attitudes\\_to\\_Data\\_and\\_AI\\_-\\_Tracker\\_Survey.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1092140/Public_Attitudes_to_Data_and_AI_-_Tracker_Survey.pdf).
- Coopamootoo, K. L. (2020, October). Usage patterns of privacy-enhancing technologies. *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*. <https://doi.org/10.1145/3372297.3423347>
- Costanza-Chock, S., Raji, I. D., & Buolamwini, J. (2022). Who audits the auditors? Recommendations from a field scan of the algorithmic auditing ecosystem [paper presentation]. ACM conference on fairness, accountability, and transparency, Seoul, Republic of Korea.
- Couchman, H. (2019). *Policing by machine*. Liberty. <https://www.libertyhumanrights.org.uk/issue/policing-by-machine/>.
- Crawford, K. (2021). *Atlas of AI*. Yale University Press.
- Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review*, 55(1), 93–128.



- Currie, M., & Podoletz, L. (2023). *Studying lived experience and automated systems: The case of universal credit*. Proceedings of the association for information science and technology.
- Data Justice Lab. (2021). *Advancing civic participation in algorithmic decision-making*. [https://datajusticelab.org/wp-content/uploads/2021/06/PublicSectorToolkit\\_english.pdf](https://datajusticelab.org/wp-content/uploads/2021/06/PublicSectorToolkit_english.pdf).
- D'Ignazio, C., & Klein, L. (2020). *Data feminism*. MIT Press.
- Duke, T. (2023). *Building responsible AI algorithms*. Apress Berkeley.
- Dumbrava, C. (2021). *Artificial Intelligence at EU Borders: Overview of applications and key issues*, European Parliament. [https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/690706/EPRS\\_IDA\(2021\)690706\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/690706/EPRS_IDA(2021)690706_EN.pdf).
- Eubanks, V. (2018). *Automating inequality*. St Martin's Press.
- European Commission. (2021). *Regulatory framework proposal on artificial intelligence*. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>.
- Evangelist, M. (2022). Narrowing racial differences in trust: How discrimination shapes trust in a racialized society. *Social Problems*, 69(4), 1109–1136. <https://doi.org/10.1093/socpro/spab011>
- Favaretto, M., Clercq, E. D., & Elger, B. S. (2019). Big data and discrimination: Perils, promises and solutions. *Journal of Big Data*, 6(12), 1–27.
- Feagin, J. (2013). Rethinking racial formation theory: A systemic racism critique. *Ethnic and Racial Studies*, 36(6), 931–960. <https://doi.org/10.1080/01419870.2012.669839>
- Fung, A. (2006). Varieties of participation in complex governance. *Public Administration Review*, 66(1), 66–75. <https://doi.org/10.1111/j.1540-6210.2006.00667.x>
- Fung, P., & Etienne, H. (2023). Confucius, cyberpunk and Mr. Science: Comparing AI ethics principles between China and the EU. *AI and Ethics*, 3(2), 505–511. <https://doi.org/10.1007/s43681-022-00180-6>
- Hall, K. (2020). *The tech firms getting their hands on NHS patient data to fight coronavirus*. Bureau of Investigative Journalism. <https://www.thebureauinvestigates.com/stories/2020-05-07/the-tech-firms-getting-their-hands-on-nhs-patient-data-to-fight-coronavirus>.
- Hern, A. (2017, July). Royal Free breached UK data law in 1.6 m patient deal with Google's DeepMind. *The Guardian*. <https://www.theguardian.com/technology/2017/jul/03/google-deepmind-16m-patient-royal-free-deal-data-protection-act>.
- Kidd, C., & Birhane, A. (2023). How AI can distort human beliefs. *Science*, 380(6651), 1222–1223. <https://doi.org/10.1126/science.adi0248>
- Li, Z. (2021). Confronting with the algorithm bias risks, if blockchain will provide new chance or challenge from data protection law perspective? *Dublin Law and Politics Review*, 2, 43–67.
- Mac Manus, S. (2021). *Dialogues about data: Building trust and unlocking the value of citizens' health and care data*. NESTA. [https://media.nesta.org.uk/documents/Data\\_Dialogues\\_Yik5MxH.pdf](https://media.nesta.org.uk/documents/Data_Dialogues_Yik5MxH.pdf).
- McIntosh, I., & Wright, S. (2019). Exploring what the notion of 'lived experience' offers for social policy analysis. *Journal of Social Policy*, 48(3), 449–467. <https://doi.org/10.1017/S0047279418000570>
- Nelson, A. (2024). The right way to regulate AI. *Foreign Affairs*. <https://www.foreignaffairs.com/united-states/right-way-regulate-artificial-intelligence-alondra-nelson>.
- Noble, S. U. (2018). *Algorithms of oppression*. New York University Press.
- O'Brien, N., Van Dael, J., Clarke, J., Gardner, C., O'Shaughnessy, J., Darzi, A., & Ghafur, S. (2022). *Addressing racial and ethnic inequities in data-driven health technologies*. Institute of Global Health Innovation. <https://doi.org/10.25561/94902>.
- O'Neil, C. (2016). *Weapons of maths destruction*. Crown Publish Group.
- Pasquale, F. (2015). *The black box society*. Harvard University Press.
- Redden, J., Dencik, L., & Warne, H. (2020). Datafied child welfare services: Unpacking politics, economics and power. *Policy Studies*, 41(5), 507–526. <https://doi.org/10.1080/01442872.2020.1724928>
- Sheehan, M., Friesen, P., Balmer, A., Cheeks, C., Davidson, S., Devereux, J., Findlay, D., Keats-Rohan, K., Lawrence, R., & Shafiq, K. (2021). Trust, trustworthiness and sharing patient data for research. *Journal of Medical Ethics*, 47(12), e26. <https://doi.org/10.1136/medethics-2019-106048>

- Taylor, D. (2022). From 'go home' vans to Windrush scandal: a timeline of UK's hostile environment. *The Guardian*. <https://www.theguardian.com/uk-news/2022/may/25/from-go-home-vans-to-windrush-scandal-a-timeline-of-uks-hostile-environment>.
- Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., & Van Moorsel, A. (2020). *The relationship between trust in AI and trustworthy machine learning technologies [paper presentation]*. Conference on fairness, accountability, and transparency, Barcelona, Spain.
- Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, 4(2), 205395171774353. <https://doi.org/10.1177/2053951717743530>
- Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review*, 2, 494–620.
- Wellcome Trust. (2011). *Community engagement: Under the microscope*. [https://wellcome.org/sites/default/files/wtvm054326\\_0.pdf](https://wellcome.org/sites/default/files/wtvm054326_0.pdf).
- Williamson, B., Komljenovic, J., & Gulson, K. (2023). *World yearbook of education 2024: Digitalisation of education in the era of algorithms, automation and artificial intelligence*. Routledge.
- Wilson, B., Twins, B., Wilkinson-Salamea, C., Raidos, D., Imafidon, K., McGarry, N., Sheni, S., & Hicks, S. (2022). *Diverse voices on data*. ClearView Research. [https://understandingpatientdata.org.uk/sites/default/files/2022-04/Diverse%20voices%20on%20Data%20-%20Main%20report\\_0.pdf](https://understandingpatientdata.org.uk/sites/default/files/2022-04/Diverse%20voices%20on%20Data%20-%20Main%20report_0.pdf).
- Wong, M. (2022). Digital society, algorithmic harm, and the pandemic response. In P. Sutoris, S. Murphy, A. Borges, & Y. Nehushtan (Eds.), *Pandemic response and the cost of lockdowns* (pp. 56–71). Routledge.
- Yin, R. K. (2008). *Case study research and applications*. Sage.