

# Just public algorithms: Mapping public engagement with the use of algorithms in UK public services

Big Data & Society  
January–March: 1–16  
© The Author(s) 2024  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20539517241235867  
journals.sagepub.com/home/bds



Helen Pallett<sup>1</sup> , Catherine Price<sup>2</sup> , Jason Chilvers<sup>1</sup>  
and Simon Burall<sup>3</sup>

## Abstract

This paper proposes and models a novel approach to public engagement with the use of algorithms in public services. Algorithms pose significant risks which need to be anticipated and mitigated through democratic governance, including public engagement. We argue that as the challenge of creating responsible algorithms within a dynamic innovation system is one that will never definitively be accomplished – and as public engagement is not singular or pre-given but is always constructed through performance and in relation to other processes and events – public engagement with algorithms needs to be conducted and conceptualised as relational, systemic, and ongoing. We use a systemic mapping approach to map and analyse 77 cases of public engagement with the use of algorithms in public services in the UK 2013–2020 and synthesise the potential benefits and risks of these approaches articulated across the cases. The mapping shows there was already a diversity of public engagement on this topic in the UK by 2020, involving a wide range of different policy areas, framings of the problem, affordances of algorithms, publics, and formats of public engagement. While many of the cases anticipate benefits from the adoption of these technologies, they also raise a range of concerns which mirror much of the critical literature and highlight how algorithmic approaches may sometimes foreclose alternative options for policy delivery. The paper concludes by considering how this approach could be adopted on an ongoing basis to ensure the responsible governance of algorithms in public services, through a ‘public engagement observatory’.

## Keywords

Algorithms, public engagement, public services, public engagement observatories

## Introduction

The adoption of algorithms in many areas of public life raises significant challenges for governance, due to issues relating to accountability, oversight, privacy, consent, and the risk of harmful outcomes such as errors, discrimination, and deepening inequalities. In response, a growing number of frameworks and approaches have emerged for ‘ethics’, ‘justice’, or ‘responsible innovation’ in relation to algorithms, many of which centre calls for and examples of in-depth public engagement.

The primary contention of this paper is that current practices of public engagement in relation to algorithms position them and their challenges as exceptional and novel, and thus fail to take advantage of three decades of relevant scholarship concerning public engagement with other emerging technologies. We identify key lessons from this large literature in science and technology studies (STS) and related fields which are relevant to the current participatory moment in algorithms research, governance and advocacy,

namely: first, that the tendency to view discrete public engagement events in isolation has made it difficult to identify broader patterns and exclusions (Chilvers and Kearnes, 2016); secondly, that the promotion of singular models of best practice, has prevented other models of public engagement such as activism or everyday actions from being taken seriously, and ignored the way that all of these models format the outcomes of participation (Chilvers et al., 2018; Mahony and Stephansen, 2016); and thirdly, that a focus on processes over outcomes has distracted from the

<sup>1</sup>School of Environmental Sciences, University of East Anglia, Norwich, UK

<sup>2</sup>School of Geography, Nottingham, UK

<sup>3</sup>Involve, London, UK

### Corresponding author:

Helen Pallett, School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK.

Email: h.pallett@uea.ac.uk



broader contexts and effects of public engagement on institutions (or the lack thereof) (Stirling, 2008; Wynne, 2006). We propose and model a novel approach to public engagement with algorithms, informed by the conceptual advances made in this literature, and which has been applied to other areas of technological innovation.

This project addresses two questions: 1) how are publics engaging with the use of algorithms in public services in the UK, and what broader themes, concerns and benefits emerge from these engagements? 2) how can we better govern the use of algorithms in UK public services? Algorithms can be defined as sets ‘*of instructions for how a computer should accomplish a particular task [... and] are used by many organizations to make decisions and allocate resources based on large datasets*’ (Donovan et al., 2018: 2). However, the term is also used to denote a broader set of approaches encompassing big data and machine learning, and to refer to instances where this automation and speedy systematic conduct of basic tasks has clear social and ethical implications and may automatically trigger further procedures (cf. Whittlestone et al., 2019).

Using a review of the academic and grey literature, supported by a stakeholder workshop with key government, civil society and private sector actors, we have mapped a diverse set of cases of public engagement with the use of algorithms in public services in the UK. From this mapping, we build the most comprehensive picture yet of how citizens are engaging with this issue and synthesise the broader hopes and concerns which are articulated through these cases.

Public services present a particularly interesting context in which to examine public perspectives and actions in relation to algorithms as it can be difficult for citizens to opt out of many of these services and their technical and data-driven requirements. Algorithm-related decisions and processes such as the determination of UK pupils’ A-level results through an OFQUAL ‘algorithm’ in 2020 and the streaming algorithm that the Home Office was revealed in 2019 to have been using for 5 years for visa processing, have proven to be both extremely consequential and controversial. The broader context of more than a decade of austerity policies also means that service providers may have unrealistic hopes about the potential of algorithms to improve services. While the empirical work for this project was completed in February 2020, the COVID-19 pandemic which began shortly after has only intensified these challenges by both accelerating the adoption of algorithms and digital technologies in public services – as well as public awareness of these approaches – and exposing gaps in provision due to austerity.

The second section of this paper reviews the relevant literature on the social implications of algorithms, the potential benefits and risks of the use of algorithms in public services, proposals to improve the regulation and governance of algorithms, and emerging approaches to improving public

engagement around algorithms and other emerging technologies. The third section describes and justifies our innovative methodology of mapping public engagement. The fourth section presents and discusses the findings of this mapping: first, describing the current state and dynamics of public engagement around the use of algorithms in public services in the UK; secondly, interpreting what these findings and the outcomes of the stakeholder workshop mean for the future responsible innovation and governance of these approaches; and thirdly, making the case for the continuing application of this mapping approach for public engagement with algorithms.

### **The effects and governance of algorithms in public services**

The use of algorithms in public services promises a range of benefits, from making the administration of public services quicker and more efficient, to removing the need for humans to perform some menial tasks, allowing greater personalization of services to different individuals, and removing the risk of human error or bias in the provision of these services (Balaram et al., 2018; The British Academy & The Royal Society, 2018). However, the adoption of algorithms across a wide range of domains of public life, from financial services to search engines has also had negative social consequences. These include the deepening and masking of forms of discrimination and stigmatization against a wide range of groups (Buolamwini and Gebru, 2018; Keyes, 2018; Madianou, 2019; Noble, 2018), the displacement of various kinds of labour from humans to machines or from one group of people to another (Taylor, 2018; The British Academy & The Royal Society, 2018), and the potential for surveillance and profiling created through the sharing and aggregation of vast data sets (Tufekci, 2019). Recent research highlights the broader political normativities and problem framings written into algorithmic systems (Birkbak and Carlsen, 2016; Crawford, 2016; Gillespie, 2014; O’Neil, 2016), the potential for decontextualized data sets to be misinterpreted (Couchman and Lemos, 2019; Dencik et al., 2019; Seaver, 2018) or contain errors (McCann and Hall, 2019), and the lack of clear accountability and transparency around these systems (Pasquale, 2016).

The negative consequences of the adoption of algorithms in public services have been particularly well-documented in the US – continuing a long legacy of justice-focussed scholarship concerning science and technology (e.g., Nelkin, 1975; Winner, 1986) – where it is argued the costs of these developments fall on already marginalized communities, who become the subjects of experiments with new algorithmic systems or bear the brunt of so-called ‘algorithmic bias’ built into algorithmic systems as a legacy of both the datasets and protocols used to train algorithms

(Donovan et al., 2018; Eubanks, 2018). In the areas of justice, social services, welfare payments and policing a key concern is that algorithmic approaches may deepen the discrimination, stigma and stereotyping faced by particular groups, resulting, for example, in low income ethnic minority communities becoming targets for predictive policing (Couchman and Lemos, 2019), or Black inmates being given lengthened prison sentences (Donovan et al., 2018).

Research in the UK has found that data aggregation processes between agencies and public service areas raise issues about privacy and informed consent, particularly for vulnerable groups like refugees (Madianou, 2019), children (Barassi, 2019), and families involved in social services programmes (Dencik et al., 2018). Although algorithmic and data-focussed projects often frame marginalized groups as their main beneficiaries, these schemes can further marginalize or inconvenience such groups through surveillance, ignoring context, limiting choice and increasing costs (Gangadharan & Niklas, 2019). Even when there are tangible benefits for marginalized groups, it is often external actors or wealthier social groups who benefit the most (Heeks and Shekhar 2019). Furthermore, a small number of private companies maintain a monopoly over citizens' digital profiles, including data sharing from public services (McCann and Hall, 2019; Shah, 2018; Sharon, 2018). In the context of healthcare provision, a key concern is that the adoption of algorithms and related technologies has the potential to widen existing health inequalities, for example by diverting funds from face-to-face services and concentrating innovative new approaches for diagnosis and care in large urban teaching hospitals (Smallman, 2019). In education, the increasing involvement of large companies in providing the educational software which schools rely on raises concerns around surveillance (Lupton and Williamson, 2017).

Given these far-reaching consequences, policymakers have been criticised for the lack of appropriate regulation and accountability which is currently in place around the use of algorithms (Donovan et al., 2018). The speed and volume of decisions made by algorithms, alongside the difficulty of disaggregating the roles played by algorithms and human decision makers in relevant processes (CDEI, 2019), poses robust challenges to governance. In common with nanotechnology (cf. Laurent, 2017) the ill-defined nature of algorithms themselves is a barrier to regulation. Algorithms and related technologies are simultaneously emergent – in that they are ill-defined, and their broader implications are still coming to light – but also already widespread through society, including in mundane contexts such as how people access information and socialise (Stahl et al., 2013). This, alongside the pervasive involvement of the private sector in innovation processes – in contrast to some other widely studied emerging technologies – and the immediacy with which ICT-based innovations can be

developed and applied, raises further regulatory and governance challenges (Jirotko et al., 2017).

The extensive literature on ethics (Balaram et al., 2018; Leslie, 2019), responsible research innovation (RRI) (Stahl and Coeckelbergh, 2016), trust (CDEI, 2021) or data justice (Dencik et al., 2022) in relation to algorithms can be summarised around a number of key themes. More technically focussed codes foreground *professional responsibility*, including ensuring the completeness and accuracy of datasets used (Leslie, 2019), guaranteeing the safety and security of systems (Burton et al., 2020), producing accompanying documentation to improve transparency and ensure responsible use (Mitchell et al., 2019), and avoiding mistakes. Approaches focussed on products and applications highlight various *social justice*-related values including human rights (Couchman and Lemos, 2019), anti-discrimination and fairness (Collett and Dillon, 2019), accessibility, and more generally promoting human values and flourishing (Stahl et al., 2021). Frameworks concerned with ensuring public *trust* in systems and applications emphasise privacy and consent (Rempel et al., 2018; Sharon, 2018), accountability (Mulgan, 2016), and transparency or explainability (Winfield and Jirotko, 2018).

This focus on ethical codes and principles follows the pattern observed around other emerging technologies such as GMOs, nanotechnology or synthetic biology, where voluntary frameworks were initially put forward as alternatives to heavy regulation by those developing new technologies. UK and European funding bodies have extended existing RRI frameworks developed around emerging technologies like synthetic biology to cover digital innovations, including algorithms. In particular, the AREA framework (Anticipate, Reflect, Engage, Act) has been institutionalised within UKRI funding streams, following Stilgoe et al.'s (2013) proposal that RRI should have four key dimensions: anticipation, reflexivity, inclusion, responsiveness. The UK's Engineering and Physical Sciences Research Council which funds researchers developing algorithms has a spin out Observatory for ethics in ICTs (ORBIT) which combines the AREA framework with Stahl and Coeckelbergh's (2016) 4Ps (People, Product, Process, Purpose) to encourage a focus on the specific challenges of ensuring RRI in ICTs, and works with researchers and the private sector. Both the AREA and 4Ps frameworks strongly emphasise public engagement as a key tool not only to achieve *engagement* and *inclusion*, but also to support other key elements of RRI and broader governance like *reflection* on and discussion of the *purposes* of innovation. Data justice frameworks, which seek to specifically highlight social justice-related concerns about algorithms and engage with the potential for such approaches to entrench inequalities and discrimination (Dencik et al., 2022), have also been accompanied by strong calls for greater public engagement.

Such calls have included proposals for the establishment of people's councils for machine learning (McQuillan, 2018), public dialogues to highlight alternative narratives of and perspectives on AI (The Royal Society, 2018) and to improve algorithmic accountability (Balaram et al., 2018), and the development of new deliberative methods to engage with the opacity of algorithms (McKelvey, 2014). Recently, civil society groups, learned societies, market research organisations and government agencies have been involved in running high profile public engagement events concerning these technologies, largely using deliberative workshop (Ada Lovelace Institute, 2021), large-scale survey (The Forum for Ethical AI, 2019) or public information campaign formats (Couchman and Lemos, 2019).

As with other emerging technologies concerns have been raised with this ethics-focussed approach to governance and associated public engagement processes. Some authors suggest AI ethics approaches are naïve in assuming that AI can simply be made fair and unbiased and criticise the 'ethics washing' of AI in the service of corporate aims, obscuring the broader intersectional injustices and power structures at play (e.g., Bui and Noble, 2020). Others criticise the 'algorithmic idealism' of trying to increase fairness through the application of mathematical techniques, instead suggesting a more radical aim of 'algorithmic reparation' for identifying the harms caused by algorithmic, and particularly machine learning systems, and as a principle for building, evaluating, adjusting, and, where necessary, omitting and eradicating machine learning systems (Davis et al., 2021). There is a specific concern that many public engagement exercises might only amount to 'participation washing', as Sloane et al. (2022) have argued in the field of machine learning. They note that publics are often engaged with machine learning applications through extractive and exploitative processes, which do not adequately consider the contexts of engagement and primarily serve the corporate ambitions of the technology developers rather than the aims and needs of engaged communities (Sloane et al., 2020). This echoes a longer running concern in the public engagement literature that such processes risk becoming a procedural or social 'fix' for the problems of technology, constructing legitimacy for institutions and decisions rather than enabling democratic governance (Irwin, 2006; Wynne 2006).

Public engagement processes around algorithms draw on well-established models of public engagement which have been in use for decades in a variety of contexts from international development, to planning, RRI and science policy making, and which have long been the subject of analysis and critique in STS and related fields. These processes share a conception of public engagement as occurring through discrete events, from which broader public views and principles can be discerned. Some of these approaches have been characterised as 'residual realist' (Chilvers and

Kearnes, 2020), meaning that they take publics and models of participation as pre-given, often by promoting a limited set of models of best practice (Chilvers et al., 2018) – usually orchestrated and framed by institutional actors – to the neglect of other 'uninvited' forms of activist participation (Leach et al., 2005) or mundane everyday actions, for example. This generally results in high quality social research on participants' views but fails to recognise the ways that these models of participation format the events themselves, with implications for what can be said, how and whether it will be considered relevant to the issue under discussion. In this kind of work, the focus on the quality of public engagement processes has detracted attention from the broader impacts and outcomes of participation, such as institutional and policy change (Pallett and Chilvers, 2013).

In response to these shortcomings, more relational approaches to public engagement have emerged which acknowledge that publics, models of engagement and issue-framings are constructed through the performance of engagement (Chilvers and Kearnes, 2016). These approaches are open to a much wider variety of formats of participation from formal deliberative workshops and surveys (Pallett and Chilvers, 2013), to forms of activism and protest (Welsh and Wynne, 2013), media communications and controversies (Marres and Moats, 2015), and mundane everyday engagements with an issue or object (Shove, 2010). These accounts acknowledge contrasting framings of the issues and objects of engagement within or between cases of engagement (Marres, 2007), rather than accepting pre-given – and often institutionally sanctioned – framings, thus creating greater opportunities to acknowledge power imbalances and political tensions (cf. Sloane et al., 2022). This approach draws attention to the partial nature of discrete public engagement processes and argues that they always need to be understood in relation to other forms of participation in wider systems (Chilvers and Kearnes, 2020).

Due to the complexity, emergence and ongoing nature of innovation systems producing algorithm-related applications, this implicit goal of producing ethical and responsible algorithms, or of creating a responsible and socially just system of innovation for these technologies, can never be definitively achieved and settled. Rather, responsible governance and innovation processes in relation to algorithms will need to be constantly practiced, monitored and debated. It follows then that forms of public engagement concerning algorithms and their uses, should not aim to definitively settle the public view on a particular technology or policy area – such as facial recognition technology, or predictive policing – but should instead recognise the multiple different scales and contexts in which these technologies pose risks and benefits to be discussed, and acknowledge that further engagement will always also be needed.

This contingency, partiality and multiplicity of public engagement, alongside the lack of a definitive fix or solution to the challenge of producing responsible, ethical and socially just algorithms for use in public services, necessitates an approach to public engagement which is **relational** – understands engagement processes in context and in relation to other processes and events –, **systemic** – looks across multiple forms and contexts of engagement for broader patterns –, and **ongoing** – allows for constant monitoring and democratic engagement with these evolving technologies and technical applications. In this paper we adapt a method developed by members of the research team to map diverse public engagement in low carbon energy transitions (Chilvers et al., 2018; Pallett et al., 2019), to provide a systemic and relational account of public engagement with the use of algorithms in public services in the UK.

## Methodology

We used a systemic mapping approach (Chilvers et al., 2021) to study public engagement with the use of algorithms in public services. This approach combines conventional systematic review methods of systematically searching the academic literature for accounts of public engagement with this topic, with additional web searches and reviews of the grey literature to identify cases of public engagement not covered in academic research. The application of the mapping method in this study comprised four main steps, as shown in Figure 1. Searches were manually carried out through academic and non-academic search engines (Web of Knowledge, Google Scholar and Google) to identify cases using search terms relating to or synonymous with ‘participation’ / ‘public engagement’, ‘public’, ‘algorithms’, ‘public services’ and ‘UK’. Given the current and emergent nature of these innovations and approaches, we identified fewer cases from the academic literature than in our previous work on low carbon transitions. Ongoing or recent examples of collectives with relevance to the topic were found in the grey literature and web searches. Additional opportunistic sampling was used to identify and follow up cases which were mentioned in the broader literature reviewed at the beginning of the project, or to follow up on suggestions of relevant cases from the project team and stakeholders, as well as media reporting of relevant cases. Most of the cases mapped were found through grey literature and opportunistic sampling.

We acknowledge that this approach does not produce a comprehensive sample – this our reason for labelling this method ‘systemic’ rather than ‘systematic’ mapping. We faced the issue of selection bias as in many systematic reviews (see de Almeida and de Goulart, 2017), as we were limited to available published or documented materials, as well as the scope of our search terms. By using search engines like Google our results are also inevitably shaped by the kinds of opaque algorithms which are the

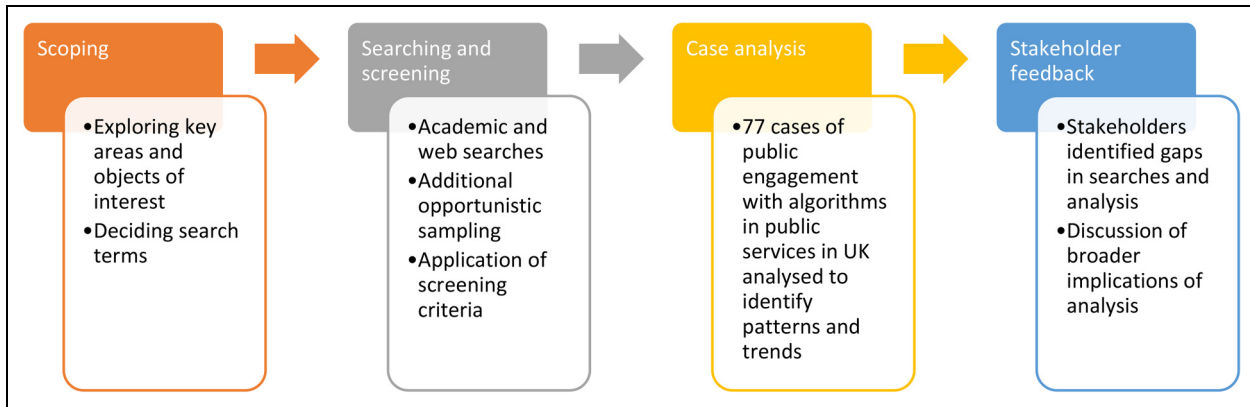
subject of this study. We have tried to mitigate for these limitations by not only including published academic literature, but also including additional cases from websites and the grey literature, where there was enough information available. We do not claim that the cases in our corpus give the full picture of public engagement with algorithms in public services in the UK during the period of study, but we do argue that this is a significantly more systemic picture than has been shown in previous accounts.

To be screened into the project corpus of relevant cases, a case had to: 1) have taken place in the UK, 2) involve some form of public engagement with the use of algorithms in public services, and 3) be sufficiently documented to allow for analysis (this means giving enough detail as to be analysed through the conceptual framework given below, and could be achieved through a significant section of an academic paper, a short report or detailed website). To achieve as diverse a sample as possible we tried to include a wide variety of algorithm-related technologies in our sample, and for example, chose to screen in cases of app-based engagement where the documentation did not specifically make clear if algorithms were being used, in order to make sure that this relevant technology and form of engagement was included in the sample.

After searching and screening, 77 cases were added to the project corpus and analysed.<sup>1</sup> The analytical framework used encompasses: a) descriptive characteristics of the cases – the public service area(s) they related to, when and where the engagement occurred, the methods used, actors orchestrating the case, and the technical affordances of algorithms the case focussed on; b) more interpretive categories concerning the relational performance of participation – the model of public engagement, the publics or participants, and the issue-framing and technical object constructed through the engagement; and c) the identification of benefits and risks of the use of algorithms in public services articulated through each engagement. This framework and the approach to coding were jointly decided upon by the project team, carried out by CP and HP, and tested by JC to ensure inter-coder reliability.

The second phase of our methodology was to run an in-person 1-day stakeholder workshop with academics, civil society representatives and policymakers with interests in algorithms in public services, held in February 2020. Workshop participants reflected on initial findings from the mapping phase which were reported through a presentation about initial findings and prepared paper templates detailing results which they worked on in small groups. These discussions identified gaps in our analysis and resulted in suggestions of new cases to include in the corpus or places to look to find cases. Participants also began to interpret and provide context to some of our findings.

In the second half of the workshop participants built on these reflections to conduct initial foresight around the use of algorithms in public services to come up with



**Figure 1.** The systemic mapping process.

recommendations for future public engagement with and governance of these approaches. This involved summarising key themes and lessons emerging from our initial analysis and making links to their existing knowledge. We gave a short presentation on emerging frameworks for mapping public engagement and public engagement observatories – including the Public Engagement Observatory of the UK Energy Research Centre (UKERC) (Chilvers et al., 2022) – and asked participants to identify key functions and activities for a potential observatory for algorithms in public services. The workshop was organised and run by SB with inputs from the rest of the project team who also contributed to group discussions throughout the day. It was recorded through notes made by CP and HP, and other outputs from group discussions including photographs, flip chart notes, post-its and inputs to paper templates produced by the project team. These outputs were analysed by the project team in a collaborative meeting following the stakeholder workshop.

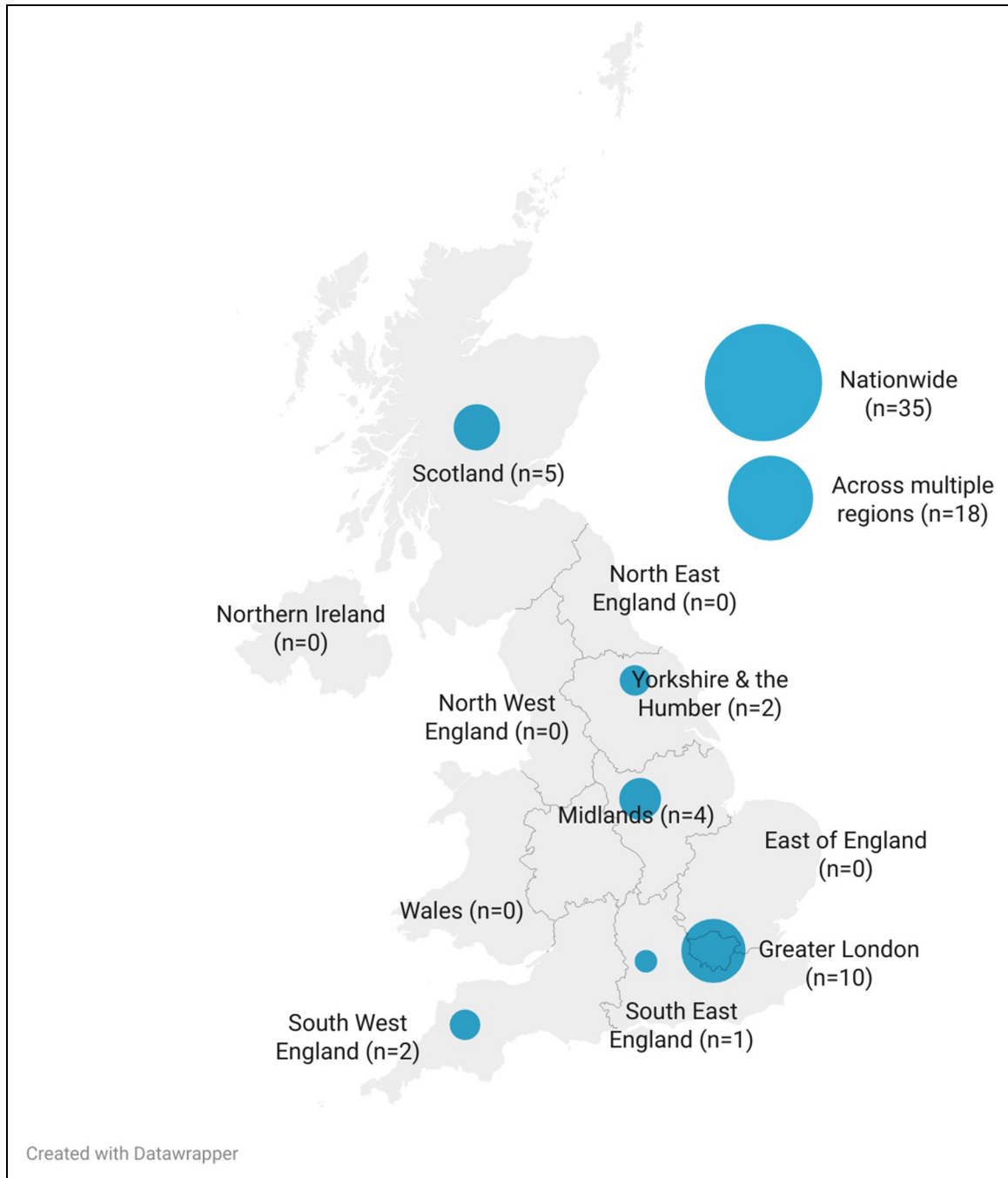
## Results and discussion

### *Public engagement with the use of algorithms in public services*

As shown in Figure 2, cases of public engagement in the corpus achieved national coverage across different regions of the UK (although none of the cases mapped specifically focussed on Northern Ireland). Most of the cases mapped attempted to or claimed to engage citizens across the UK (either nationwide or across multiple regions). A significant number of cases were based in London, for example introducing and monitoring new algorithmic monitoring on Transport for London services, or trialling new approaches to service delivery in London teaching hospitals. In a minority of cases, attempts were made, especially through deliberative engagement, to engage with citizens in more marginalised or diverse sets of locations in England, Scotland and Wales.

Cases were found and mapped from 2013 onwards, but as Figure 3 shows numbers of cases generally saw a more substantial increase from 2016 onwards. The emergence of significant controversies around facial recognition technology and risk assessment analytics provided a focus for many cases 2018–20, whereas earlier cases tended to be concerned with data sharing and collection. Bodies related to UKRI (such as the public dialogue body Sciencewise and the medical research consortium the Farr Institute), independent research bodies (such as the Nuffield Foundation and the Wellcome Trust), and professional bodies (such as the Royal Statistical Society) were significant in orchestrating these earlier engagements which were often concerned with more hypothetical uses of algorithms in public services. More recent cases were characterised by heavier involvement of Government departments and campaigning organisations (such as Liberty and Big Brother Watch), and generally focussed on the ways algorithms were already being used in public services. Cases orchestrated through academic studies appear in the dataset from 2018 onwards.

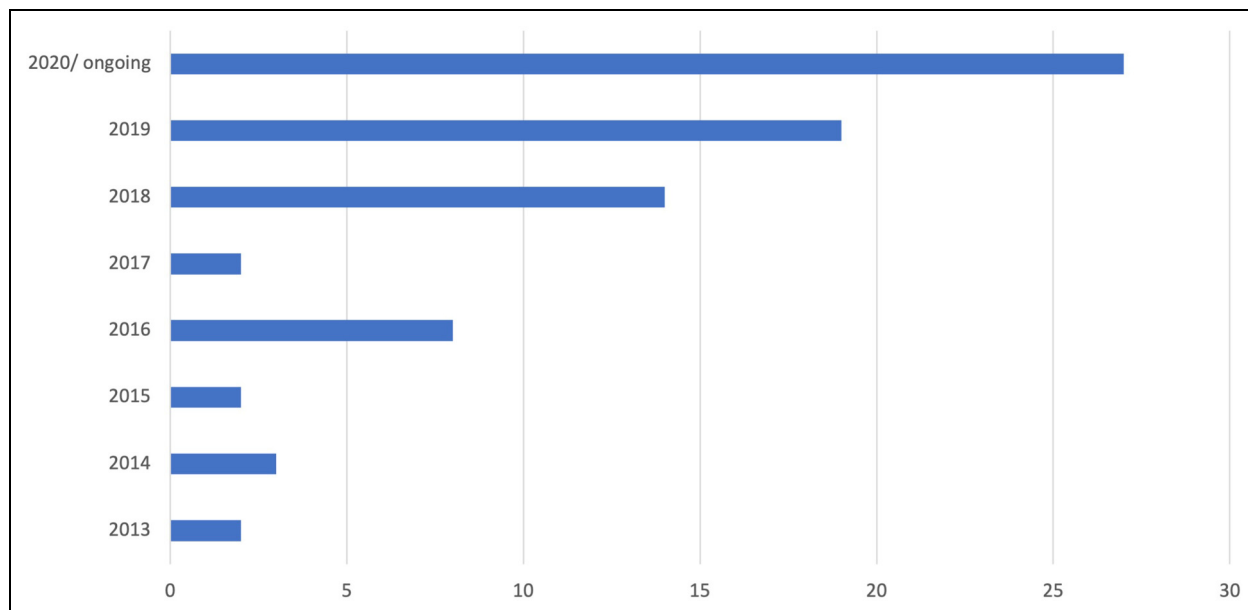
Table 1 shows how we categorised the cases according to different public service areas associated either with a particular government department or agency. Health and social care was the public service area which accounted for the largest number of cases mapped, with 33 cases overall. This is the area where we found the most cases of mundane everyday engagement with algorithms in public services, for example through apps and chatbots. In this area, there have been discussions and public engagement ongoing for some time, accounting for a lot of the earlier cases. Around social care specifically, media investigations were significant for raising awareness or forming the centre of controversies around the use of algorithms in public services as modes of engagement. Specialist organisations related to healthcare have been significant in orchestrating engagement, especially the NHS and The Wellcome Trust (including its Understanding Patient Data programme). Much of the public engagement in this area uses deliberative workshops and surveys as the primary models of engagement.



**Figure 2.** The geographical distribution of cases of public engagement with the use of algorithms in UK public services, as identified in the mapping.

Policing was the other significant public service area in our mapping (19 cases), characterised by a large range of very different kinds of public engagement – from protests to information campaigns, live trials of new technologies, and court cases – orchestrated by very different kinds of organisations – including police bodies, professional legal bodies, campaign groups, and academics.

Immigration and justice were the focus of only 6 and 4 cases respectively, and these were generally more polarised and controversial. There was lots of activity by campaign groups and professional associations around these areas, including information campaigns and legal challenges. Our mapping found a similar pattern for education (5 cases) with specialist campaign organisations like ‘digitaldefendme’ orchestrating protests and mass



**Figure 3.** Number of cases of public engagement with algorithms in public services by start date.

**Table 1.** Cases by public service area.

Public service area	Number of cases
Health and social care	33
Policing	19
FigureTransport	7
Immigration	6
General	6
Education	5
Justice	4
Local services	4
Environment	2
Welfare	2
Broadcasting	2
Regulation of private sector	1
Defence	1
Financial Services	1
Work	1

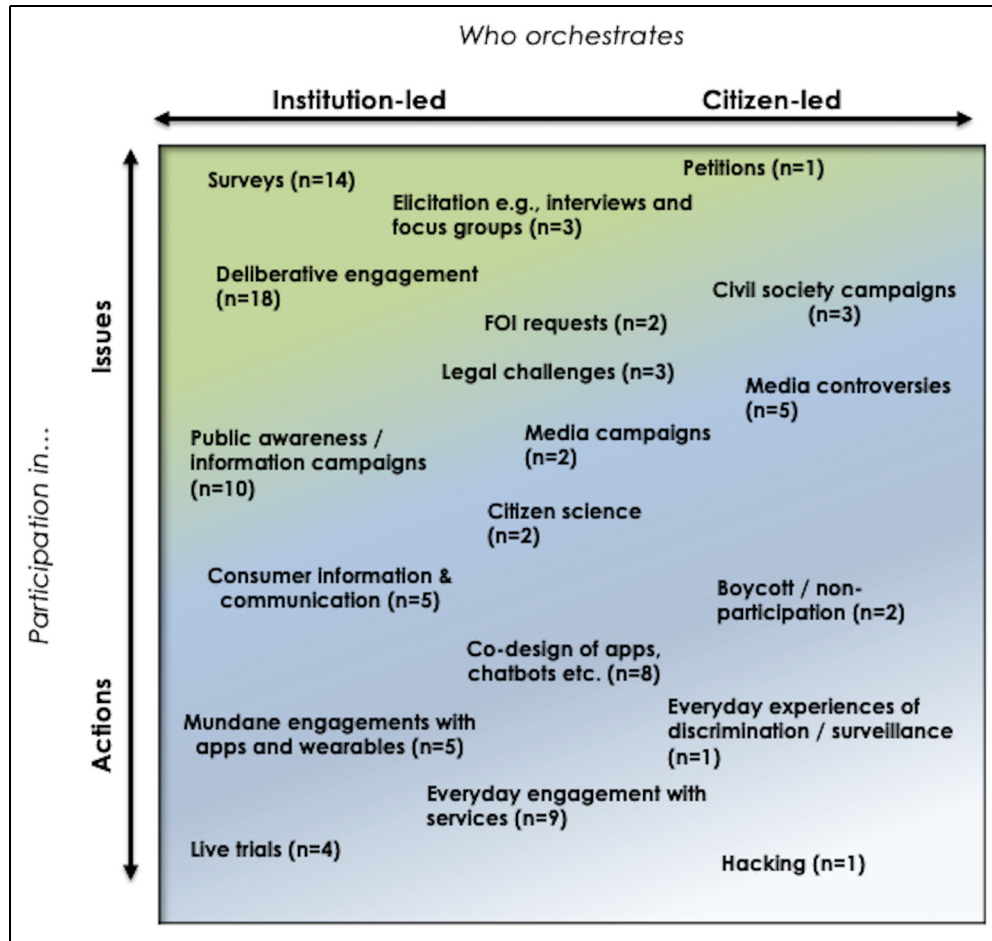
non-compliance around data collection practices. Transport (7 cases), environment (2 cases), broadcasting (2 cases) and local services (4 cases) were generally domains with more forms of everyday public engagement and also domains where fewer concerns have thus far been raised around the use of algorithms in public services. While defence (1 case) and welfare (2 cases) have been more controversial areas in academic work on algorithms, we did not pick up on many examples of public engagement in these areas. Six cases in the mapping concerned the use of algorithms in public services more generally and did not identify a more specific public service area as a focus, for example, Ipsos MORI's 2016 public dialogue on the ethics of data science in government.

A major finding of this mapping is the sheer number of existing cases of public engagement around this topic, and the diversity of forms of public engagement being carried out in diverse institutional settings. Furthermore, many of the cases mapped included more than one model of engagement. This shows, citizens do have the capacities to engage with these technologies, and to raise important issues about their implementation, even in the absence of detailed technical understanding.

Deliberative forms of engagement (18 cases) – especially public dialogues (5 cases) and citizens juries (5 cases) – were commonly used models for public engagement with the use of algorithms in public services, as were surveys (14 cases). This is particularly notable in cases orchestrated by independent research bodies or UKRI-related bodies. Public awareness campaigns, media campaigns and general communication approaches were also significant (17 cases), encompassing many cases orchestrated by charities and campaign groups. Mundane forms of engagement with technologies in use like apps, wearables and chatbots linked to public service provision were another significant area in the mapping (17 cases). New forms of engagement mapped here which have not been found in similar mappings conducted for citizen engagement around climate change and low carbon transitions (see Chilvers et al., 2021), included FOI requests, live trials (mainly of facial recognition technology) and boycotts of or non-compliance with data sharing.

In Figure 4, we use the heuristic framework for mapping diversities of public engagement developed by Chilvers et al. (2021) to plot these forms of engagement based on two continua according to whether they are more





**Figure 4.** A mapping of diverse public engagements with the use of algorithms in UK public services, 2013–2020.

institution- or citizen-led, and whether they are more action- or issue-oriented. These categories were arrived at through qualitative analysis of written materials related to each case, which enabled us to identify whether a case was being driven more by a formal institution such as a government body or research funder or citizens, and whether it was more focused on public talk and information provision (and therefore more issue-oriented) or more oriented towards engagements with material objects or concerning material practices (and therefore more action-oriented). This diagram is presented as a heuristic only, acknowledging that these categories are themselves co-produced and not absolute.

This reveals that most of the cases mapped are more institution-led, though more citizen-led approaches seem to be emergent and have become more common in recent years. There is a fairly even distribution between cases more concerned with issues and those more concerned with action. Some of the more action-oriented forms of engagement were quite mundane, including engagement with chatbots and apps designed for service users and

unwitting participation in trials of technologies like facial recognition technology in policing. More emergent citizen-led and action-oriented forms of engagement included the co-design of new apps and chatbots with users, deliberate non-participation in apparently compulsory algorithmic systems such as databases for aggregating education records, and the use of face paints by activists to ‘hack’ facial recognition systems.

Unsurprisingly, given the dominance of institutionally orchestrated forms of public engagement, as well as a significant number of more mundane engagements with the issue, a dominant construction of the participants involved in 30 of the cases mapped was as ‘service users’ – including some more specific subgroups such as pregnant people / new parents and patients. A more general construction of participants was as ‘affected citizens’ (13 cases), indicating that they had been selected, recruited or assembled because of their pre-existing experiences of and perspectives on the problem in question. This included people who had been wrongly accused of criminality and gang membership through facial recognition technology and risk assessment

analytics, or those involved as subjects in the notorious ‘Troubled families initiative’. In 7 cases, participants were specifically engaged as ‘unaffected citizens’ as their lack of prior experience of or stated perspectives on the problem in question were considered to make them ideal participants in processes such as citizens juries and public dialogues. ‘Interested citizens’ were the main participants of 18 cases mapped, indicating that participants self-selected due to interest in the problem under discussion, particularly as targets of information campaigns or activist initiatives. As found in previous mapping work, the notion of an ‘aggregate population’, which is a demographically representative cross-section of the population as a whole was a powerful way of constructing participants in high-profile cases (6 cases), particularly surveys of public understanding and attitudes.

A further construction of citizens in this mapping is the imagination of ‘the general public’ (4 cases) – particularly present in cases concerning policing, justice and defence – as the public which is being protected and safeguarded through the use of algorithms in public services, or which is potentially subjected to trials of facial recognition technology and risk assessment analytics without active or knowing participation.

Data collection, sharing and use were the technical affordances of algorithms which formed the focus of 19 cases mapped, and this interest was constant throughout the period of the mapping (2013–2020). More recent intense interest in the technical affordance of facial recognition technology was the focus of 13 cases. These cases covered a range of different engagements from more activist engagement to more formal deliberative processes. Apps, dashboards, and wearables (13 cases), as well as chatbots (6 cases), were both the focus and means of public engagement in several cases – especially covering more mundane engagements and ‘service user’ publics. Predictive analytics (4 cases) and risk assessment analytics (7 cases) were also a significant technical focus of cases and tended to bring out significantly more controversy and evidence of public concerns. These affordances were a particular focus of a lot of activist engagement, information campaigns and academic inquiry. More specific technical applications such as biometrics, drones, automated vehicles and virtual systems were also the focus of a small number of cases.

### *Responsible innovation of algorithms in public services*

Public engagement processes serve an important role in anticipating the future consequences of emerging technologies. Therefore, the mapping conducted for this paper can also be used as a basis for thinking about the responsible future innovation of algorithms in public services.

In 29 of the cases mapped, participants articulated a sense that algorithms would lead to improved services, such as better diagnoses of illnesses, better care and treatment, reduced congestion, improved public health outcomes or reduced crime. These ideas were often part of the initial framing of the engagement – often by those with interests in encouraging the use of algorithms in public services such as government bodies or technology companies. Such benefits were often invoked quite vaguely without clear causal mechanisms expressed. A 2016 Ipsos MORI public dialogue on the ethics of data science in government concluded a clear public benefit needs to be established for each context in which such approaches are to be applied but found few examples of this in practice. The adoption of cyber kiosks – desktop software allowing police officers to view the contents of a mobile phone or tablet – by Police Scotland provides a further illustration of these dynamics. The public engagement process orchestrated through Police Scotland’s formal trial of this approach frames the outcomes of cyber kiosks as protecting the most vulnerable. However, a case of public engagement with the cyber kiosks orchestrated by the campaign group Open Rights Group and the charity Privacy International, instead frames the primary outcome as infringements of human rights.

Greater efficiency (10 cases) and cost savings through better allocation of resources (8 cases) were the other most commonly articulated benefits of the use of algorithms in public services in the cases mapped. A key discussion point in the stakeholder workshop was that these kinds of benefits might be mainly enjoyed by public service providers themselves, rather than being benefits that are directly felt by citizens. However, direct benefits to service users were articulated in a small number of cases, including improved customer service (6 cases), better decision making and prediction of needs (6 cases), improved accuracy and consistency (3 cases), and reductions in bias and inequality (2 cases). Specific aspects of public service provision where algorithms could provide immediate benefits were seen as being in information provision (6 cases) and enabling 24/7 service provision (3 cases).

The potential harms of the use of algorithms in public services raised in public engagement processes varied according to the applications or aspects of algorithms and related technologies which the cases of engagement focussed on, as summarised in Table 2. In the context of engagements concerned with data collection, storage and sharing the primary harms identified related to issues of privacy, informed consent, data security, confidentiality and anonymity. In the context of algorithms being applied in public service contexts the primary harms highlighted concerned issues of discrimination, bias and inequality, as well as recognising the potential for malign uses of algorithms, mistakes and unintended consequences. Concerns were raised in some cases around broader issues of

**Table 2.** Potential harms of the use of algorithms in public services.

Potential harms	Number of cases
<b>Data collection, storage and sharing</b>	
Infringement of privacy	18
Sharing or storage of data without a subject's knowledge	12
Lack of informed consent	7
Failures in data security	6
Unnecessary surveillance	5
Lack of awareness of data use sharing	3
Use of data for other purposes	3
Commercial organisations profiting from public data	3
Not ensuring confidentiality and anonymity	2
<b>Applications of algorithms to public services</b>	
Discrimination	15
Inaccuracies and mistakes	13
Bias	11
Undermining or compromising human rights	7
Deliberate misuses, harm and abuse	7
Exacerbation of existing inequalities	4
Hacking	2
Lack of empathy	2
Unintended consequences	1
Wasted time	1
<b>Governance and regulation</b>	
Lack of transparency	8
Lack of accountability	7
Lack of oversight, regulation and governance	6
Lack of attention to fairness and inequality	4
<b>Service provision and alternatives</b>	
Substitution for other (face to face) services	2
Complicity in damaging policy agendas	1
Meeting service providers' needs rather than service users' needs	1

governance and regulation, including lack of transparency, accountability and attention to fairness. A much smaller number of cases articulated concerns about the use of algorithms in public services foreclosing alternative models of service provision which might place more emphasis on face-to-face contact, particularly with the backdrop of austerity policies and cost-cutting. For example, the Care Data Futures Dialogue run by the independent organisation Doteveryone concluded that algorithmic approaches were potentially foreclosing other models of care, and prioritising service provider's needs over those of service users. In some cases, algorithms were seen as problematic because they were considered complicit in damaging policy agendas like the 'Hostile Environment' or the Troubled Families Programme.

Concerns about bias and discrimination were very significant and came up across a range of different public service domains including healthcare, justice, policing, immigration and social care. Concerns about surveillance

and human rights were mainly related to policing and immigration. The potential harms caused by inaccuracies and mistakes were explored across a range of different public service areas, such as mistaken identities in facial recognition technology uses for policing, overly long sentences given through the justice system, and misdiagnoses of illnesses.

An overall finding of the mapping is that differently framed and orchestrated cases of public engagement tend to articulate very different public views about the outcomes of the use of algorithms in public services. This counsel against an overreliance on interpreting the outputs of single cases of public engagement. Recent engagements around the use of facial recognition technologies (FRT) in UK public services are illustrative of this point. More institutionally orchestrated public engagement, such as the Metropolitan Police's live trial of these technologies, framed public protection and upholding public order as the main outcomes of adopting FRT. An online survey carried out by the Ada Lovelace Institute – an organisation which was explicitly set up to carry out public engagement around these issues and which is funded by the long-established Nuffield Council for Bioethics – found significant concerns around consent, trust, surveillance, bias, discrimination and mistakes, alongside these potential benefits. Similar concerns were raised by media reporting and controversy around the topic, and by an open-source map of applications of FRT in the UK created by an investigative journalist. Engagements orchestrated by more explicitly activist and campaigning organisations such as Liberty and Big Brother Watch, raised similar concerns but also highlighted the potentials for human rights violations. An ethnographic research project conducted by legal researchers at the University of Essex also highlighted human rights concerns (Fussey and Murray, 2019).

*Towards an observatory for algorithms in public services?*

This paper has shown the value of public engagement to produce meaningful insights into the challenges of responsibly governing algorithms in public services. Furthermore, mapping diverse cases of public engagement gives a more comprehensive picture of the multiple framings and affordances of the objects of interest, as well as public perspectives on potential benefits and risks. This provides more comprehensive and nuanced evidence from which to inform decision-making and practice than the conventional reliance on one-off, high-profile, and institutionally orchestrated forms of public engagement. Furthermore, as Broomfield and Reutter (2022) have argued, a focus on institution-led engagement alone on this topic is likely to involve at best a limited role for

citizens, often understood just as ‘users’ and engaged in a tokenistic way.

Here we join others in the public engagement literature concerning the challenges of governing other emerging technologies such as gene editing (Burall, 2018; Jasanoff and Hurlburt, 2018) and low carbon innovations (Chilvers et al., 2021) in arguing for new infrastructures and institutions – such as observatories – to support the ongoing mapping of public engagement and to fulfill a number of related functions concerning governance and decision-making. Observatories recently established by research organisations, internationally in the case of the Global Observatory for Genome Editing (Saha et al., 2018) and nationally in the case of the UKERC Public Engagement Observatory (Chilvers et al. 2022), offer frameworks that can inform the development of similar entities for the responsible governance of algorithms and AI.

Concerning accelerating global developments in research and application of gene editing technologies, Saha et al. (2018: 742) argued for the need for a forum to support “*more sustained, iterative, and inclusive revisiting at the global level of key questions surrounding [these] technologies*”. In contrast to a public dialogue, consultation or opinion-survey approach to integrating societal concerns into governance and decision-making, the observatory models put forward emphasise the value of drawing on existing experience and wisdom on a topic (Saha et al., 2018) and taking advantage of existing networks and communities through which citizens are already engaging with the issue (Burall, 2018; Chilvers et al., 2022), deliberately seeking to bring in and examine perspectives on and framings of the issue which have been overlooked or dismissed by other bodies (Saha et al., 2018), and foregrounding societal questions around and framings of the problem (Saha et al., 2018).

In the case of algorithms, singular public engagement processes are likely to focus on individual technologies and affordances and may focus on isolated parts of a distributed innovation system. Therefore, they are unlikely to give insights into connections between different instances of public engagement and patterns at a systemic level such as the impacts of austerity or the COVID-19 pandemic which emerge through an observatory and mapping-focussed approaches (Chilvers et al., 2022). The adoption of algorithms in the context of public services is accelerating and they are also evolving in use, therefore there is a need for constant monitoring and discussion over their governance going forward. The COVID-19 pandemic brought to the fore a range of issues related to sharing of health data with private companies, surveillance potential of tracking contacts, movements and infection risk, as well as broader issues related to the treatment of immigrants. This is leading to an explosion in new cases of engagement around the use of algorithms in public services which raise new challenges and concerns for policy actors to

pay attention to, further justifying the need for a public engagement observatory in this area.

In addition to continually monitoring and mapping public engagement with the use of algorithms in public services, participants in our stakeholder workshop pointed to several additional useful activities which could be carried out by an observatory, which parallel suggestions from the existing literature. Supporting coordination between different bodies orchestrating public engagement processes around the issue to avoid the duplication of efforts was an activity which many workshop participants saw as valuable. Similarly, Burall (2018) argues that observatory bodies should view themselves as active nodes in existing networks taking on the role of facilitating and revealing links between existing organisations and communities. More ambitiously, Hurlbut et al. (2018) also argue for the need for observatory-style bodies to enable cooperation between countries around the governance of emerging technologies.

Workshop participants also saw a role for an observatory in designing and orchestrating public engagement which deliberately seeks to engage underrepresented issues, communities or forms of engagement which have been identified through mappings. This parallels Burall’s (2018) argument that observatories should seek to engage publics through a range of different mechanisms – from social media to formal deliberation – and Saha et al.’s (2018) call for a focus on important questions around emerging technologies which would otherwise be neglected.

Producing synthetic resources which make visible the findings of mappings and draw broader lessons from existing examples of public engagement was another activity put forward by workshop participants, as evident in existing observatory approaches (Chilvers et al., 2022). This reinforces two of the major observatory activities proposed by Saha et al. (2018) to first gather and make visible a range of ethical and policy responses to a given emerging technology, and second to provide substantive analysis of emerging conceptual developments, tensions and areas of consensus around these technologies. Workshop participants also saw a useful role for an observatory in creating space for reflection about what the outcomes of public engagement might mean for governance and regulation of algorithms, also evoking Saha et al.’s (2018) proposal that an observatory should serve as a forum for convening periodic discussions on important questions.

Echoing enduring arguments made through STS, public engagement and anticipatory governance literatures about the value of public engagement with emerging technologies, a final observatory activity proposed by workshop participants was to actively contribute to the anticipation of possible futures of algorithmic approaches in UK public services, as well as the possible futures of public engagement itself (cf. Chilvers & Kearnes, 2020).

## Conclusions

This paper has shown that citizens have the capacities to engage meaningfully with the governance of algorithms in public services and other contexts. They are already doing so in the UK through a variety of different forms of engagement and around different framings of the issue and technical affordances of the application of algorithms. The 77 cases of public engagement mapped here do articulate some benefits of the use of algorithms in public services, but they also point out serious concerns which are not being adequately addressed through narrow, technically-focussed risk assessment and governance approaches. The analysis offered in this paper also shows the importance of asking broader questions about whether these algorithm-based approaches are always needed, and where the benefits will be felt.

Public concerns and hopes about the adoption of algorithms in UK public services are not adequately captured by abstract lists of core principles. Rather they point to more political and justice-based issues which require open discussion and debate, and genuine ongoing engagement, especially when we look beyond institutionally orchestrated and issue-focussed cases of engagement. Decisions about applications of algorithms need to be assessed against alternatives, rather than simply weighing up potential costs and benefits of one form of action.

Moving forward, it is crucial to continue to foster and respond to meaningful public engagement with the use of algorithms in public services, and other contexts. But academic and policy approaches to public engagement need to be relational, systemic and ongoing in order to address the complex governance challenges presented by these technologies which are emergent, impossible for human actors to entirely oversee, and continually innovating in a distributed manner.

This paper has demonstrated and developed one possible methodological approach for mapping diverse cases of public engagement in the context of the use of algorithms in public services. We do not claim that the mapping offered here is comprehensive or is the only or best way to approach public engagement as relational, systemic and ongoing. However, this approach does produce an account of public engagement which is more comprehensive than the conventional reliance on singular public engagement events. It also opens up to diverse formats of participation, including those that are citizen-led and reveal alternative public concerns which can challenge institutionally sanctioned processes or extractive and exploitative processes associated with ‘participation washing’. There continues to be a need to develop and refine new methods for mapping public engagement to operationalise this new conceptualisation of public engagement and its role in innovation governance, and to make it possible to continually map engagement in a resource efficient manner. We particularly see a role for

the development of new digital methods in this area to more quickly spot and map emerging engagements, for example, as well as an important role for qualitative and ethnographic social science methods to help gain better understandings of forms of engagement which have been under documented and to explore the contexts and interconnections of the cases mapped.

Underlying the analysis presented here including 25 cases that highlight concerns over governance and regulation (Table 2), and much of the material cited, is the recognition of an urgent need to develop a basis for appropriate regulation and oversight of the use of algorithms in public services. This needs to go beyond asking how algorithms can be made ‘ethical’ or attempting to prevent the most obvious examples of misuse of algorithms. Rather there are deeper issues to be addressed, such as ensuring algorithms are being adopted to address genuine needs and problems, that they form part of broader policies and structures which reflect democratic values and seek to improve the lives of public service users, and that the datasets from which consequential decisions are made are contestable, transparent and accountable. Crucially, given both the nature of innovation processes of algorithms and emerging arguments about the conceptualisation of public engagement, these regulation and oversight structures also need to allow for ongoing monitoring of innovation and public engagement processes, be open to challenge and contestation, and therefore also incorporate revisions rather than upholding a rigid structure.

## Acknowledgements

The authors would like to acknowledge the contributions of participants from across civil society and the public and private sectors who attended our stakeholder workshop in February 2020 to give feedback on initial project findings and contribute to discussions about the design of a public engagement observatory to ensure responsible governance of algorithms.

## Declaration of conflicting interests


The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Engineering and Physical Sciences Research Council, (grant number EP/R044929/1).

## ORCID iDs

Helen Pallett  <https://orcid.org/0000-0001-5014-6356>

Catherine Price  <https://orcid.org/0000-0003-1846-5407>

## Note

1. Full descriptions and further coding details for each of the 77 cases in the corpus are provided in the open access dataset available here: [https://figshare.com/articles/dataset/Just\\_Public\\_Algorithms\\_project\\_database\\_of\\_cases/14438777](https://figshare.com/articles/dataset/Just_Public_Algorithms_project_database_of_cases/14438777).

## References

- Ada Lovelace Institute (2021) *The Citizens' Biometrics Council: Recommendations and findings of a public deliberation on biometrics technology, policy and governance*. Report. London, UK: Ada Lovelace Institute. March.
- Balaram B, Greenham T and Leonard J (2018). Engaging citizens in the ethical use of AI for automated decision-making. Report, The royal society for the encouragement of arts, manufactures and commerce (RSA), London UK, May.
- Barassi V (2019) Datafied citizens in the age of coerced digital participation. *Sociological Research Online* 24(3): 414–429.
- Birkbak A and Carlsen HB (2016) The public and its algorithms: Comparing and experimenting with calculated publics. In: Amoores L and Piotukh V (eds) *Algorithmic Life: Calculative Devices in the age of big Data*. London: Routledge, pp. 21–34.
- Broomfield H and Reutter L (2022) In search of the citizen in the datafication of public administration. *Big Data & Society* 9(1).
- Bui M and Noble SU (2020) We're missing a moral framework of justice in artificial intelligence. In: Dubber M, Pasquale F and Das S (eds) *The Oxford Handbook of Ethics of AI*. Oxford: Oxford University Press, 162–179.
- Buolamwini J and Gebru T (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research* 81: 1–15.
- Burall S (2018) Don't wait for an outcry about gene editing. *Nature* 555: 438–439.
- Burton S, Habli I, Lawton T, et al. (2020) Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective. *Artificial Intelligence* 279: 103201.
- Centre for Data Ethics and Innovation (CDEI) (2019). *Interim report: Review into bias in algorithmic decision-making*. Available at: <https://www.gov.uk/government/publications/interim-reports-from-the-centre-for-data-ethics-and-innovation/interim-report-review-into-bias-in-algorithmic-decision-making> (accessed 30/01/2024).
- Centre for Data Ethics and Innovation (CDEI) (2021). *COVID-19 Repository & Public Attitudes: 2020 in Review*. Available at: [https://assets.publishing.service.gov.uk/media/60473e258fa8f577c0fae7c6/CDEI\\_COVID19\\_Repository\\_and\\_Public\\_Attitudes.pdf](https://assets.publishing.service.gov.uk/media/60473e258fa8f577c0fae7c6/CDEI_COVID19_Repository_and_Public_Attitudes.pdf) (accessed 30/01/2024).
- Chilvers J, Bellamy R, Pallett H, et al. (2021) A systemic approach to mapping participation with low-carbon energy transitions. *Nature Energy* 6(3): 250–259.
- Chilvers J and Kearnes M (2016) Participation in the making: Rethinking public engagement in co-productionist terms. In: Chilvers J and Kearnes M (eds) *Remaking Participation: Science, Environment and Emergent Publics*. London: Routledge, pp. 31–63.
- Chilvers J and Kearnes M (2020) Remaking participation in science and democracy. *Science, Technology & Human Values* 45(3): 347–380.
- Chilvers J, Pallett H and Hargreaves T (2018) Ecologies of participation in socio-technical change: The case of energy system transitions. *Energy Research and Social Science* 42(April): 199–210.
- Chilvers J, Pallett H, Hargreaves T, et al. (2022) *An Observatory for Public Engagement with Energy and Climate Change*. Report. London: UK Energy Research Centre.
- Collett C and Dillon S (2019) *AI and Gender Four Proposals for Future Research*. Report, Levehulme Centre for the Future of Intelligence, Cambridge UK.
- Couchman H and Lemos AP (2019) *Policing by Machine*. Report, Liberty, London UK, February.
- Crawford K (2016) Can an algorithm be agonistic? Ten scenes about living in calculated publics. *Science, Technology & Human Values* 41(1): 77–92.
- Davis J, Williams A and Yang MW (2021) Algorithmic reparation. *Big Data & Society* 8(2): 1–12.
- de Almeida CPB and de Goulart BNG (2017) How to avoid bias in systematic reviews of observational studies. *Revista CEFAC* 19: 551–555.
- Dencik L, Hintz A, Redden J, et al. (2018) *Data Scores as Governance: Investigating uses of citizen scoring in public services*. Report. Cardiff, UK: Cardiff University.
- Dencik L, Hintz A, Redden J, et al. (2022) *Data Justice*. London: Sage.
- Dencik L, Redden J, Hintz A, et al. (2019) The 'golden view': Data-driven governance in the scoring society. *Internet Policy Review* 8(2): 1–24.
- Donovan J, Caplan R, Matthews J, et al. (2018) *Algorithmic Accountability: A Primer*. Report. Washington: Data & Society. April.
- Eubanks V (2018) *Automating Inequality*. New York: St Martin's Press.
- Forum for Ethical AI (2019) *Democratising decisions about technology: A toolkit*. Report, The royal society for the encouragement of arts, manufactures and commerce (RSA), London UK, October.
- Fussey P and Murray D (2019) *Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology*. Report University of Essex Human Rights Centre, UK, July.
- Gangadharan SP and Niklas J (2019) Decentering technology in discourse on discrimination. *Information, Communication & Society* 22(7): 882–899.
- Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski P and Foot K (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge: MIT Press, pp. 167–194.
- Heeks R and Shekhar S (2019) Datafication, development and marginalised urban communities: An applied data justice framework. *Information, Communication & Society* 22(7): 992–1011.
- Hurlbut JB, Jasanoff S and Saha K, et al. (2018) Building capacity for a global genome editing observatory: conceptual challenges. *Trends in Biotechnology* 36: 639–641.
- Irwin A (2006) The politics of talk: Coming to terms with the 'new' scientific governance. *Social Studies of Science* 36(2): 299–320.
- Jasanoff S and Hurlbut B (2018) A global observatory for gene editing. *Nature* 555: 435–437.

- Jirotko M, Grimpe B, Stahl B, et al. (2017) Responsible research and innovation in the digital age. *Communications of the ACM* 60(5): 62–68.
- Keyes O (2018) The misgendering machines. *Proceedings of the ACM on Human-Computer Interaction* 2(CSCW): 1–22.
- Laurent B (2017) *Democratic Experiments: Problematizing Nanotechnology and Democracy in Europe and the United States*. Cambridge: MIT Press.
- Leach M, Scoones I and Wynne B (2005) *Science and Citizens: Globalization and the challenge of engagement*. London: Zed Books.
- Leslie D (2019) *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. Report*. London: The Alan Turing Institute.
- Lupton D and Williamson B (2017) The datafied child: The data-veillance of children and implications for their rights. *New Media and Society* 19(5): 780–794.
- Madianou M (2019) The biometric assemblage: Surveillance, experimentation, profit, and the measuring of refugee bodies. *Television & New Media* 20(6): 581–599.
- Mahony N and Stephansen H (2016) The frontiers of participatory public engagement. *European Journal of Cultural Studies* 19(6): 583–597.
- Marres N (2007) The issues deserve more credit: Pragmatist contributions to the study of public involvement in controversy. *Social Studies of Science* 37(5): 759–780.
- Marres N and Moats D (2015) Mapping controversies with social media: The case for symmetry. *Social Media + Society* 1(2): 205630511560417.
- McCann D and Hall M (2019). *Digital Self-Control*. Report, New Economics Foundation, London UK, March.
- McKelvey F (2014) Algorithmic media need democratic methods: Why publics matter. *Canadian Journal of Communication* 39(4): 597–613.
- McQuillan D (2018) People’s councils for ethical machine learning. *Social Media and Society* 4(2).
- Mitchell M, Wu S, Zaldivar A, et al. (2019) Model cards for model reporting. FAT\* 2019 - Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency, 220–229.
- Mulgan G (2016) *A machine intelligence commission for the UK: how to grow informed public trust and maximise the positive impact of smart machines*. Report, Nesta, London UK, February.
- Nelkin D (1975) The political impact of technical expertise. *Social Studies of Science* 5(1): 35–54.
- Noble S (2018) *Algorithms of oppression*. New York: NYU Press.
- O’Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. London: Penguin.
- Pallett H and Chilvers J (2013) A decade of learning about publics, participation and climate change: Institutionalising reflexivity? *Environment and Planning A* 45(5): 1162–1183.
- Pallett H, Chilvers J and Hargreaves T (2019) Mapping participation: A systematic analysis of diverse public participation in the UK energy system. *Environment and Planning E: Nature and Space* 2(3): 590–616.
- Pasquale F (2016) *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press.
- Rempel ES, Barnett J and Durrant H (2018) Public engagement with UK government data science: Propositions from a literature review of public engagement on new technologies. *Government Information Quarterly* 35(4): 569–578.
- Saha K, Hurlbut JB, Jasanoff S, et al. (2018) Building capacity for a global genome editing observatory: Institutional design. *Trends in Biotechnology* 36(8): 741–743.
- Seaver N (2018) What should an anthropology of algorithms do? *Cultural Anthropology* 33(3): 375–385.
- Shah H (2018) Algorithmic accountability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376: 20170362.
- Sharon T (2018) When digital health meets digital capitalism, how many common goods are at stake? *Big Data & Society* 5(2): 1–12.
- Shove E (2010) Beyond the ABC: Climate change policy and theories of social change. *Environment and Planning A* 42: 1273–1285.
- Sloane M, Moss E, Awomolo O, et al. (2022) Participation is not a design fix for machine learning. In EAAMO ‘22: Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization 1: 1–6.
- Smallman M (2019) Policies designed for drugs won’t work for AI. *Nature* 567(7746): 7–7.
- Stahl BC, Andreou A, Brey P, et al. (2021) Artificial intelligence for human flourishing – beyond principles for machine learning. *Journal of Business Research* 124(November 2020): 374–388.
- Stahl BC and Coeckelbergh M (2016) Ethics of healthcare robotics: Towards responsible research and innovation. *Robotics and Autonomous Systems* 86: 152–161.
- Stahl BC, Jirotko M and Eden G (2013) Responsible research and innovation in information and communication technology: Identifying and engaging with the ethical implications of ICTs. In: Owen R, Bessant J and Heintz M (eds), *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*. Chichester: John Wiley & Sons, pp. 199–218.
- Stilgoe J, Owen R and Macnaghten P (2013) Developing a framework for responsible innovation. *Research Policy* 42(9): 1568–1580.
- Stirling A (2008) Opening up” and “closing down”: Power, participation and pluralism in the social appraisal of technology. *Science Technology & Human Values* 33(2): 262–294.
- Taylor A (2018) The automation charade. *Logic: A Magazine about Technology*, 1 August.5.
- The British Academy & The Royal Society (2018) *The Impact of Artificial Intelligence on Work*. Report, London UK, September.
- The Royal Society (2018) *Portrayals and perceptions of AI and why they matter*. Report, London UK, November.
- Tufekci Z (2019) Machines shouldn’t have to spy on us to learn. *Wired*, 25 March.
- Welsh I and Wynne B (2013) Science, scientism and imaginaries of publics in the UK: Passive objects, incipient threats. *Science as Culture* 22(4): 540–566.
- Whittlestone J, Nyrupe R and Alexandrova A, et al. (2019) *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research*. London, UK: Nuffield Foundation.

- Winfield AFT and Jirotko M (2018) Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376: 20180085.
- Winner L (1986) *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. Chicago: University of Chicago Press.
- Wynne B (2006) Public engagement as a means of restoring public trust in science - hitting the notes, but missing the music? *Community Genetics* 9(3): 211–220.