Centring the Margins: Mapping AI Systems as Systems of Power

Garfield Benjamin
University of Cambridge
Professional and Continuing Education
garfield.benjamin@pace.cam.ac.uk

Abstract

This paper introduces a method of critically mapping AI as an assemblage of social relations. This approach is rooted in the principles of centring the margins and highlighting power structures. Viewing an AI or algorithmic system as a wide-reaching network of social relations, including their impacts on different groups and contexts, enables consideration of how AI is embedded within different discourses and domains of power while emphasising the impact on those most affected. The paper provides a discussion of the critical framing of the project, the principles, processes and templates for mapping AI in this way, and three examples of algorithmic systems in public sector contexts that have been discontinued, are in use now or are being proposed for future use. A discussion of uses and limitations is provided, situating the method beyond a descriptive or analytical tool towards a critical approach to identifying locations for intervention in harmful or unjust uses of algorithmic systems.

Extended version (with full-page figures) —
https://papers.ssrn.com/abstract=5381429

Introduction

Anthropologist and feminist science and technology studies scholar Lucy Suchman aptly describes the current discourses around AI as a narrative of "uncontroversial thingness" (Suchman 2023). Seeing an AI system as a "thing" is an effective allegory to make complex systems tangible and perhaps understandable as a specific collection of functions or technical objects, particularly outside of academic discussions. But when this allegorical shorthand spreads far and wide in the press, education, corporate marketing and even back into academic critique, it carries significant risks of obscuring how AI systems are embedded within vast sets of social relations. Beyond a thing, beyond hype, beyond function, in this article we instead begin from an understanding of AI as relations and narratives of power.

A global foresight report from the United Nations Environment Program made "a stark reminder of the interconnectedness and fragility of our systems in the 21st Century" (UNEP 2024) not only relating to planetary health but also human wellbeing. The report reflects Pasek et al.'s focus not only on economic or even environmental factors but also the social structures surrounding the energy usage of AI (Pasek et al. 2023). Partly, this attention involves the merging of technology companies and energy companies, seen more recently in Amazon, Microsoft and Google renting or buying power plants to fuel the long-term needs of their data centres and planned AI infrastructures (Blunt 2024). Partly, it relates to the limits of, for example, Moore's Law, leading to an "outsourcing" of compute power to maintain the industry and government myths of perpetual miniaturisation, perpetual increase in efficiency and perpetual economic growth (Pasek et al. 2023). These changes in turn feed into the expansion of AI power into new - and often less visible - sectors. Across environmental and social issues, it has become incredibly difficult to identify not only where power lies around an AI system but also where interventions can be made to mitigate harms, offer resistance, and support those most affected. What methods, then, can be used to map AI as social systems with the potential for change?

This project builds on the intersection of science and technology studies (specifically critical algorithm studies and critical data studies) and critical cartography (Ash, Kitchin, and Leszczynski 2019; Baumann et al. 2024; Lin 2015; Carraro 2023). This includes the map or atlas of AI (Crawford and Joler 2018; Crawford 2021) which added to the literature highlighting the infrastructural needs and material reach of producing specific technological artefacts (Kitchin 2014), and the following environmental harms, labour exploitation and inequalities. Other work has focused on mapping technical and corporate relations and ecosystems that situate, for example, platforms and apps (van Angeren, Alves, and Jansen 2016; Dieter et al. 2019) in their social, economic and political contexts, and outlined methods for participatory and narrative methods of mapping in relation to data practices, communities and critical literacies (Feigenbaum and Alamalhodaei 2020; Aleman and Martinez 2024). The present article combines aspects of these approaches while shifting the focus to centre those most affected, emphasising social relations. This is motivated by the need to highlight power and inequality as well as the role of discursive, social and political structures in shaping AI.

What is offered is a critical method for grappling with hidden, sprawling systems of power and influence. Beyond a compendium, it is one way of addressing these often ob-

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

fuscated, abstract or discursive problems in concrete ways. It opens specific discussions about specific AI systems for use with relevant critical and practitioner communities across academic, policy, civil and technology sectors. We argue against completeness and instead encourage the identifying of gaps in knowledge (by, for example, scavenging or participatory approaches), and emphasise creating space for critical discussion that leaves open question marks where, for example, proprietary training datasets preclude direct audit or analysis. Filling in these gaps is useful, but it is also useful to locate such gaps and situate how decisions were made to obscure such information. This follows Suchman's emphasis on situating oneself in social relations, towards locating accountability in specific places and relations of power (Suchman 2002). The aim is therefore not to present a laying of the land but offer social and discursive framings of AI in order to provoke critique and direction: towards activism and locating points for intervention that centre the needs of those most affected by AI systems.

The paper starts with a critical framing of the method, including engaging with debates on toolkits, assemblages, and cartographies, as well as the Black feminist theories that inform the core understanding of AI that shapes the proposed method. Then the mapping method itself is outlined through its core principles, the process of developing a map, and templates for the framework as well as a generalised AI map taking into account common aspects of algorithmic systems. Three example maps of systems in the UK public sector context are then presented, demonstrating: an algorithmic decision-making system that was rejected, the UK Office of Qualifications exams algorithm from 2019; a system in use, the Department for Work and Pensions fraud and risk assessment system; and a proposed use of a particular type of AI system, Police Scotland's potential use of live facial recognition technology. Following these examples, a discussion of the potential for the method for a range of contexts and purposes is offered, along with its limitations and specificities. Finally, the paper concludes with a restatement of its purpose and pathways for use and future research.

Framing

A driver for the project of this paper is to centre those most affected by AI systems, and this is most often those who are already marginalised by societies in various ways. As mapping the margins formed the basis of intersectionality at its first enunciation by Kimberlé Crenshaw (Crenshaw 1991), this will be a guiding principle when assessing who is affected and the multiple dimensions that may affect particular individuals or groups. Black feminisms therefore frame the discussion of power and pathways to resistance, specifically by taking seriously bell hooks's push to centre the margins (hooks 1984) and Patricia Hill Collins's critique of the matrix of domination (Collins 2002) within the context of assessing the impacts of AI systems. More broadly, the project is informed by principles of situated knowledge (Haraway 1988), epistemic justice (Fricker 2007), and highlighting the margins (Rodríguez 2017) when constructing knowledge. These theoretical underpinnings set out the approach to centring the margins and addressing the power structures embedded in specific configurations of AI as social relations.

As a driver for contestation, refusal and alternative modes of conceptualising AI assemblages, we are motivated by framings such as Ngozi Okidegbe's application of afrofuturist of thinking across past, present and future to untether the design, development and deployment of algorithms from inequitable logics and discourses with a more transformative vision for algorithms that "can endow Black and other marginalized people with a voice to reform, reimagine, or dismantle a system that has traditionally subordinated their values, needs, and interests" (Okidegbe 2022). This builds on Safiya Noble's Black feminist critique of "algorithms of oppression" (Noble 2018) in a move towards more justice-focused "epistemic practices around algorithms" (Okidegbe 2024). These approaches also align with the focus across Ruha Benjamin's body of work on imagining alternative futures outside of current technology narratives. Partly, this is a matter of agency over visibility. Following Akwugo Emejulu and Francesca Sobande's (Emejulu and Sobande 2019) work on existence as resistance (combining Black feminism and Afrofeminism in the context of European digital cultures), we can problematise the simultaneous hypervisibility (for example, in data) and invisibility (in design and decision-making processes) of marginalised groups. Mapping does not only make visible otherwise marginalised points on the assemblage that forms an AI system. It can instead involve "a commitment to privileging – rather than 'equally including'" the voices, experiences and needs of marginalised groups (Tanksley and Estrada 2022). This is an essential step towards the transformative potential of a critical map that brings in knowledge, power and action. What this means in practice is centring those most affected and using the mapping process to make visible the asymmetric power relations that would otherwise silence those individuals or groups.

However, mapping is not neutral, particularly towards groups historically marginalised by race, geography, gender or other indices of oppression. When constructing an approach to mapping, we must therefore consider critiques of cartographic thinking and alternative understandings of cartographies as situated forms of knowledge. Mapping has a colonial history, and it is important to retain a reflexive critical approach when using mapping as a tool. Maps tell stories, weaving narratives of power into the fabric of understanding. This can offer alternative approaches to territory that aid in generating ways of making new worlds (Escobar 2018), specifically in terms of critiquing AI systems as "planetary assemblages of coloniality" (Hung 2024). However, it can also impose and entrench dominant discourses of power, and mapping should therefore be confronted with situated contestations of colonial conceptions of territory, particularly in, for example, activist responses to AI and data infrastructures (Lehuedé 2022). To critique, and to constitute alternative – even radical – arrangements of power requires alternative ways of mapping. This project builds on feminist situational mapping (Clarke 2003) and feminist geographies and spatialities of intersectional thinking through the spatial formations of interlocking violence(s) (Mollett and Faria 2018) in order to provide such critique and alternative. It is indebted to Black feminist cartographies of struggle and

critiques of mapping (McKittrick 2006, 2020) as well as cartographies of resistance across gender, class and geography (Kannabiran 2006) and anti-racist and anti-colonial mapping that emphasise social construction of spatial orientations involving critiques of geographic imaginaries of the past to in connection with (and to better inform) more collective and constructive imaginations for the future (Kobayashi 2014, 2019). The project builds on approaches such as data feminism's repurposing of information to highlight and challenge power (D'Ignazio and Klein 2020), queer geographies that critically emphasise identity, agency, history, structure and power in assemblages (Kinkaid 2020), and creative approaches rooted in imagination (Harmon 2004) to make visible existing power structures and create space for collective action.

Maps in Sara Ahmed's queer phenomenology act as technologies with a normalising role (Ahmed 2020). For maps, these "technologies" relate to navigation and so normalise specific bodies (self and others) and specific directions (orientation in relation to objects and others). Undertaking specific forms of mapping AI, then, considers these processes of orientation and normalisation of bodies and directions that are embedded by a given AI system. This might be power relations that invoke resistance or refusal, which are given directionality and a particular relation to different bodies through the process of mapping (labour exploitation and unionised resistance, for example). It is important, therefore, to consider the particular ways in which the map is not the territory, for the map encourages openness, connection and reworking, an active and ongoing process rather than a state of certainty and competence (Deleuze and Guattari 2004). However, Eden Kinkaid identifies in shallow conceptions of assemblage a tendency towards flux and transformation that reiterates existing power dynamics and obscures asymmetries and hierarchies (Kinkaid 2020), reminding us to ensure that power relations and asymmetries are made visible.

Mapping processes such as the method proposed in this paper follow the aim of resistance and refusal that identify and challenge the "knowledge projects" that shape what kinds of harm and change are valid or possible, driven by a commitment to the social "not just as sites of problem-solving, but more explicitly as a place where directions to answers and solutions to harms actually exist" (Ganesh and Moss 2022). Our aim is to highlight the constraints of mainstream discourses and the violence entailed by the interaction between discourse, power and technical infrastructures (Hoffmann 2019). We return to Katherine McKittrick's work on Black feminist geographies as processes that contest and rupture existing landscapes of power, being concerned with "the terrain of political struggle itself, or where the imperative of a perspective of struggle takes place" (McKittrick 2006). These intersectional and radical forms of mapping are expressive demands. Building approaches from the perspective of different forms of marginalisation generates "a cartography of governance that forces an official reckoning of a new way of seeing" (Kannabiran 2006). Starting from these different intersecting critiques and practices, we employ mapping here in a way that confronts established conceptions of how we understand and spatialise AI and algorithmic technologies as well as providing new ways of imagining alternative configurations of power that centre those currently most marginalised.

Map

The mapping method presented here spans critical humanist and conflict approaches to understanding AI in society, with the goals being both "to describe and problematize assumptions in order to identify potential for change" and "to identify and modify patterns of domination", through approaches of "insight through critical analysis" and "liberation through structural analysis" (Sovacool and Hess 2017). We turn now to the translation of the conceptual framing of our mapping method into a practical, visual approach that embodies this framing. The construction of the method is therefore undertaken with a commitment to the following guiding principles:

Centring the margins is the guiding principle (hooks 1984) of how the map is constructed. Rather than a diagrammatic representation of technical structures that prioritise artefacts as a starting point, we propose that those who are affected by an algorithmic or AI system should be at the centre of our understanding. Conceptually, this means engaging with principles of community and representation, building on work that highlights the different forms of marginalisation at work and elevating the perspectives of those who are affected (Milan and Treré 2021; Milan, Treré, and Masiero 2021). Practically, and visually, we take this imperative literally by building our mapping method around a centre of those most affected, especially those most harmed by a given AI system.

Plural perspectives are essential to a social analysis of a technical system, and the commitment to Black feminisms and queer phenomenologies of this method is rooted in an opposition to universalising conceptions of knowledge (Haraway 1988; Fricker 2007) and technical systems (Milan and Treré 2019). This goes hand in hand with a rejection of completeness in our understanding of a map as an ongoing process rather than a fixed state. It requires us to make intellectual space for error, complexity and self-critical reflection. When constructing the map visually, this reminds us to tease out relations of labour and decision-making as well as to find ways of representing different contexts of social relations, creating space for alternative approaches, and encouraging creative uses of the framework.

Situating within power and discourse follows on from these first two principles. The aim of conceiving of AI as a social system in relation to how certain groups are affected leads us towards a focus on power relations. The map is not flat, but an unequal topology where power asymmetries have directionality. Simultaneously, the justification and shaping of these power structures are supported by discourses and contexts which are themselves often unequal. Both conceptually and visually, then, we situate the map within Patricia Hill Collins's domains of power (Collins 2002). This matrix of domination is made up of four domains: *Structural power* is how oppression and power asymmetries are organised, largely at the societal level. This may include global

supply chains or geopolitical considerations, or the overarching frameworks that shape how decisions are made within a given context. Resistance in this area tends to require transforming social systems. *Disciplinary power* manages oppression through social institutions and other related organisations. It particularly relates to issues of surveillance and bureaucracy, and resistance tends to come from within the relevant organisations. *Cultural power* is also called hegemonic power and relates to ideas, discourses and norms, and the ways they are spread. These forces shape the limits of what decisions are (or can be) made, and support structural and disciplinary domains. Resistance can come from self-definition and collective consciousness, and seeking the epistemic justice of empowering alternative ways of thinking. *Interpersonal power* is the everyday experiences of oppression by specific individuals and groups, perpetuating oppression through the collection of small acts that reiterate oppression. Resistance here is often tactical and contextual interventions in specific relations or interactions. These domains of power divides the map not by geographically spatial contexts but adds a spatial dimension to power relations. By locating a given person, object, process, institution or narrative in a specific place on the map, we are invited to consider where else they may sit in relation to the other elements. We are therefore not only visualising social entities or forces in an AI system but also making visible and giving directionality to the relations between them. In doing so, the map can help uncover and contest the intersections and interactions of different forms and contexts of power that shape relations and lived experience, and fundamentally define the ways that an AI system affects different people and groups in different ways. Constructing a map in this way is a distinctly political act that challenges stable power structures and dominant narratives.

Purpose and direction intellectually, methodologically, politically and practically follow on from and underpin an inconvenient and critical process. Given the way that mapping AI can be a seemingly endless and sprawling process, to retain clarity it will always be necessary to collapse or condense certain elements. For example, while environmental harms should be a major concern for any AI development, they may not be the core focus as part of a policy discussion around inequalities. This does not mean that such concerns should be ignored, but that a mapping process will inevitably have certain bounds and that these should be made consciously within the context of its purpose. Doing so allows the mapping process to engage with critical discussions in relevant ways while still leaving space to open up wider conversations and make visible previously ignored impacts. Acknowledging the purpose of the map also gives a clearer direction for its construction and use. One of the challenges in addressing algorithmic harms and wrongs is identifying the places where effective interventions can be made. The potential avenues of action (as we demonstrate in our examples below) can cross different sectors and different scales. Often it requires a combination of pressures and practices in order to mitigate immediate harms and especially in order to make larger systemic changes or shift discourses and imaginaries around what algorithms and AI can and should do. Therefore, alongside different harms, we can look at different methods of repairing algorithms. For example, approaches to co-constituting terms with users, addressing friction, enabling refusal mechanisms, complaint and reporting mechanisms, and disclosure-centred mediation (Rakova, Shelby, and Ma 2023). This approach of repair can be undertaken even when a particular system is seemingly "not broken" (Pop Stefanija and Pierson 2024), either to prepare in advance to mitigate potential harms or to look past potentially overspecialised responses to highly specific harms in order to situate and tackle structural wrongs within broader aims of reparative justice (Davis, Williams, and Yang 2021; Doyle et al. 2024). The mapping process builds on these approaches towards social justice and challenging existing power structures that enable unjust uses of AI.

Across these principles is a commitment to the underlying drive towards critical and structural analysis in order to provide insights for relevant groups - researchers, developers, policymakers, publics - and to establish avenues for liberation. The map allows for points of tension, resistance and refusal to be identified, and can help focus efforts in relevant places. The visual construction of the mapping method follows these aims.

Constructing the Map

Based on the above framing and principles, the map is constructed as follows:

- Centre those most affected: the map is orientated around those most affected by the system, with those normally most marginalised closer to the centre;
- Around this is the technical layer: the different technical objects sit in relation to one another, and the people or groups they directly affect;
- Connected is the labour layer: around the technical objects we position those whose work goes into making those objects and systems possible, whether in a technical sense (broadly construed) or those who do the work of operations, developing policy, running services, organise and resist, or other roles connected to the development and deployment of the system;
- Further out is the influence layer: the outer ring is occupied by those who make decisions, the organisations who effect different levels of control, as well as the structures and discourses that shape how an AI system is developed or used.

This creates the core construction of the mapping frameworkTo show the different ways in which these different objects, roles, actors, organisations and discourses shape a given AI system, we can also assign them to the domains of power. We separate different functions and roles into Patricia Hill Collins's (Collins 2002), four domains of power: *Structural power*; *Organisational power*; *Interpersonal power*; *Cultural power (hegemonic power)*. Overlapping these two schema leads to the main template, shown as a legend in Fig.1. We considered fully decentring the technical objects and placing them outside the circle of the map, but this risked on one hand surrounding the system with technol-

ogy in a deterministic way, and on the other hand made connecting those affected more visually confusing. Instead, we opted to retain the technical object as an "interface" layer of sorts, but removed the hard boundaries between layers. The map is porous, and there is much overlap between roles. For example, those in the labour category could be more or less exploited, more or less empowered. A low-paid casualised data worker in the Global South with no corporate contract and few employment rights is in a very different position from a lead designer at a big tech company in Silicon Valley with a high salary, stock options and other benefits. Power is higher the further out from the centre, which echoes the centring of the margins, the structural nature of oppression, and the role that social discourses play in shaping power relations. We also suggest that items can be further coded by category: we propose person/group, object/practice, institution/organisation and concept/narrative; colours are used to identify technological objects, unknowns and, perhaps most importantly, points of intervention. Others may adapt these to their needs; however, this was the most effective way of representing these relations visually, acknowledging the spatiality and directions of influence and power.

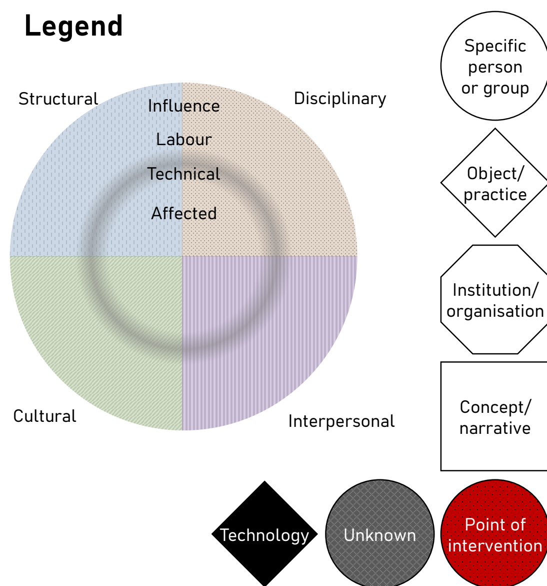

Figure 1: Legend

Further aspects of power and critique can be embedded within the visualisation of the map. For example, the lines can be used to show missing, incomplete or ineffective links (dashed) or connections with stronger influence (double). Similarly, arrow heads give directionality to show the impact of power. This differs from existing maps that trace the flow of data and instead emphasises the impact of power relations and decision-making or influence. Nodes on the map can be coloured to show particular points of interest. In our examples, black is used for technical objects, red (dotted) shows points of contestation or potential intervention which are particularly useful for purposes of resistance and refusal, and grey (cross-hatched) show gaps or obscured knowledge for further investigation or to demonstrate the epistemic politics surrounding transparency and accountability within the assemblage.

Process

In order to populate the map (as is shown in examples below) for a specific AI system, we recommend an approach to gathering data based in queer and Black feminist practices of scavenging (Niang 2024). This methodology is entangled, open, and non-extractive. It focuses on attention, repair and worldmaking. There is a reflexivity and positionality inherent to this process, building on feminist notions of situatednesss and partiality (Haraway 1988). For example, we acknowledge in our examples a UK focus based on our own geographical location, an outsider status in relation to official documentation and reliance on publicly available materials, the facticity of the authors (UK-based, queer, non-binary, white, educated), as well as the privileges that come with an academic position combined with a commitment to addressing power inequalities in AI systems at a structural level. Other instances generated by other groups would take on a different narrative, building different worlds. This is an integral part of the mapping process: while it is a useful visualisation of concrete relations, it is more importantly an expression of a particular story of power around a given AI system. We would encourage a collaborative process for the use of the map, as we have been doing in educational contexts with students from a wider range of backgrounds, contexts and lived experiences. The researcher here is a facilitator for engagement and discussion, providing their own expertise in sociotechnical and discursive conceptions of AI, rather than a keeper of a prescriptive process.

On a practical level, this process uses what is already there rather than extracting new data from marginalised groups. This means following trails of public records, recent moves towards algorithmic transparency registers, freedom of information requests (Clifton-Sprigg, James, and Vujić 2020; Walby and Luscombe 2020), company records, technical documentation (including, where appropriate, code), policy, legal documentation (such as privacy policies and terms and conditions for third parties), journalism, media and other sources such as bias trackers (Smith and Rustagi 2021). This approach to scavenging sits alongside Tanksley's "Black feminist archival bricolage", which adds to uncovering the narratives and perspectives in the construction of the map with a commitment to making visible and centering the impact of algorithms on marginalised groups (Tanksley 2023). This can involve embedding stories within subjective experiences of technology, going beyond reactive critique towards actively imagining possible alternatives (Shaw, Toliver, and Tanksley 2024), echoing the importance of subjectivity in the mapping process. This methodology acknowledges involvement and positionality, generates expansive questions rather than driving for completeness, focuses on social rather than technical relations, and sees the gathering of data as being integrated with theory, mapping and analysis. The scavenging process replaces data collection with a reflexive pro-

cess of constantly situating and analysing power relations in ways that will inform the creation of the maps. It also encourages participatory approaches (Aleman and Martinez 2024) that enable the mapping method to be used alongside community-building or education. Different contexts may require different methods, and as we demonstrate below, there is value in identifying where information is obscured or missing, as a way to help determine future research and future action.

Examples

We now offer three examples of the mapping method applied to three specific systems. All are taken from the UK public sector context, but span systems that were used and then rejected, are in use now, and are being proposed. The first is the 2020 Office of Qualifications (OfQual) exams algorithm, the second is the Department for Work and Pensions (DWP) risk and fraud assessment algorithm, and the third is Police Scotland's potential use of facial recognition. Each is presented with a specific purpose to demonstrate the different ways the method can be used.

Rejected: OfQual Exams Algorithm

The 2020 A-Level exam results algorithm developed by the Office of Qualifications (OfQual, the UK qualification regulator) was developed in response to the Covid-19 pandemic and the inability of students to sit exams, confronting teacher assessed grades with a political fear of grade inflation. On instruction from government ministers, OfQual designed an algorithm that used teacher assessed grades to rank students within their year group and then assigned grades based on that rank in comparison to the same ranked students averaged over previous years from that school. This resulted in large numbers of students having their grades reduced, and an inequality in grades based on location and previous successes of the schools. In response to this, student protests, legal challenges, and media coverage all combined to shift public perception and put pressure on the government (Benjamin 2022), leading to the algorithm being withdrawn and students receiving the higher of algorithmically or teacher assigned grades. The map in Fig.2 is completed using data scavenged from OfQual's own reporting (OfQual 2020), press and other sources that engage with personal experiences of those affected (BBC 2020), as well as academic analysis (Kelly 2021; Benjamin 2022). The map includes named individuals and organisations, such as those bringing legal action or particular students whose grades were lowered and who appeared in news coverage. We show the distribution of affected people and groups across interpersonal (everyday), structural and disciplinary domains. Key components include the protest practices that instigated change and influenced wider discourses, shifting algorithmic imaginaries in various sectors. There is therefore a strong focus on the cultural or hegemonic domain as it relates to changing ideas and a disruption of established assumptions about the deployment of algorithms. Other highlights are the different elements of the structural and disciplinary border constituted by the different individuals and departments from government implicated in the incident and targeted by protest and formal legal review mechanisms. This example demonstrates the mapping method as a historical analytical tool for assessing not only algorithmic harms and wrongs but modes of intervention, redress and resistance. The map emphasises points of refusal and modes of collective action. This purpose - highlighting successful modes of "decomputerisation" (Penn 2021) guides the design of this particular map, and its insights can provide tactical examples (Heemsbergen, Treré, and Pereira 2022) for activists and advocates seeking to address future harmful algorithms.

In Use: DWP Fraud Risk Assessment

The second example demonstrates the use of the mapping method to analyse an existing algorithm currently in use. The UK Department of Work and Pensions (DWP) uses an algorithm for fraud and risk assessment of welfare claims, specifically for those applying for a cash advance within the Universal Credit benefits system. The system is trained on historical data of personal claims with parameters set to detect the risk of fraud. New individual claims are compared to this dataset. Notably, it explicitly includes the use of protected characteristic data or proxies thereof. The algorithm flags cases for human review, leading to potential rejection of claims and possible criminal investigation. The mapping method acts as a critical tool not only for identifying harms and wrongs, and the power structures and discourses they are situated in, but also for identifying blank spaces. In this particular case, there have been barriers to transparency mechanisms that create a very limited capacity to actually assess bias. Data for the map in Fig.3 is scavenged from DWP reports (of Work and Pensions 2023), press coverage which itself collects data from government and external sources (Booth 2023), and a later freedom of information request (Jones 2024). Absent are detailed impact assessments or the inclusion in the Algorithmic Transparency Recording Hub, for example. The press source includes perspectives from other government regulators, such as the National Audit Office, who provided objections and warnings about the bias inherent in the system (which were promptly ignored or overridden by DWP and government minister narratives and priorities, while the Freedom of Information request was met with a largely redacted response that identified the fact that bias was occurring but not what that bias was or who was actually being affected. These gaps raise significant additional questions and the mapping method offers a way to visualise and confront these structural barriers. There are epistemic injustices here whereby critics, researchers and activists can identify that there are problems but are not able to access enough information to provide evidence that could support potential victims of the system. The visual presentation of the map includes double-weighted lines to show the most dominant line of influence from discourse to DWP, and highlights the points where information is missing or obscured. Key components here are the unknowns: the redacted freedom of information report and the unspecific groups impacted. Future engagement with wider publics might help fill in these gaps. The components in power are largely structural and disciplinary, in the mechanisms of state that define and implement unequal systems of social welfare supported

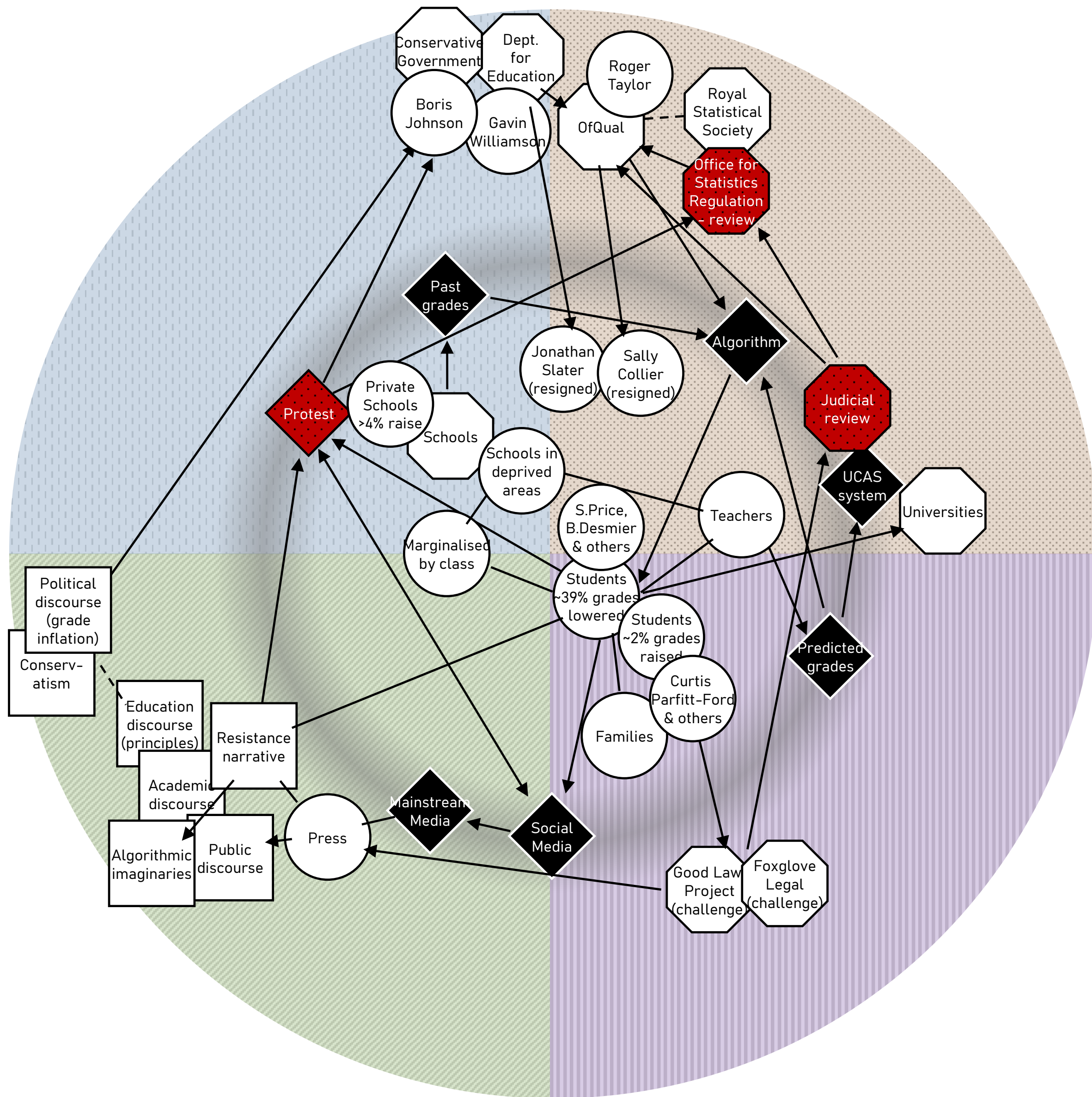

Figure 2: Map of the decision-making that led to the OfQual exams algorithm and the resistance practices that emerged to counter it

by these biased AI systems. The technical objects therefore act as a displacing interface between the applicant and those responsible for making the structural decisions. The purpose of this map is to make visible those places where more information is needed and expose the power structures that enact such obfuscation. This opens space for further investigation (academic, legal, journalistic) that can engage with alternative methods of filling in the gaps left by official documentation and redacted freedom of information requests. (Fiorella 2019). It thereby offers critical analysis of the connections between these structural and epistemic inequalities and the technological and organisational systems that entrench them, as well as identifying points to apply political pressure for transparency and accountability towards a more detailed assessment of the impacts of this particular algorithm.

Proposed: Police Scotland Facial Recognition

The final example we present demonstrates the map as a forward looking method for discerning prospective impacts and potential areas of intervention. As the focus on technologies as social power relations engages deeply with discourses and decision-making, it is important to focus on critically analysing these processes as they happen. This example shows the discussions around Police Scotland's proposed use(s) of live facial recognition technologies. Following on

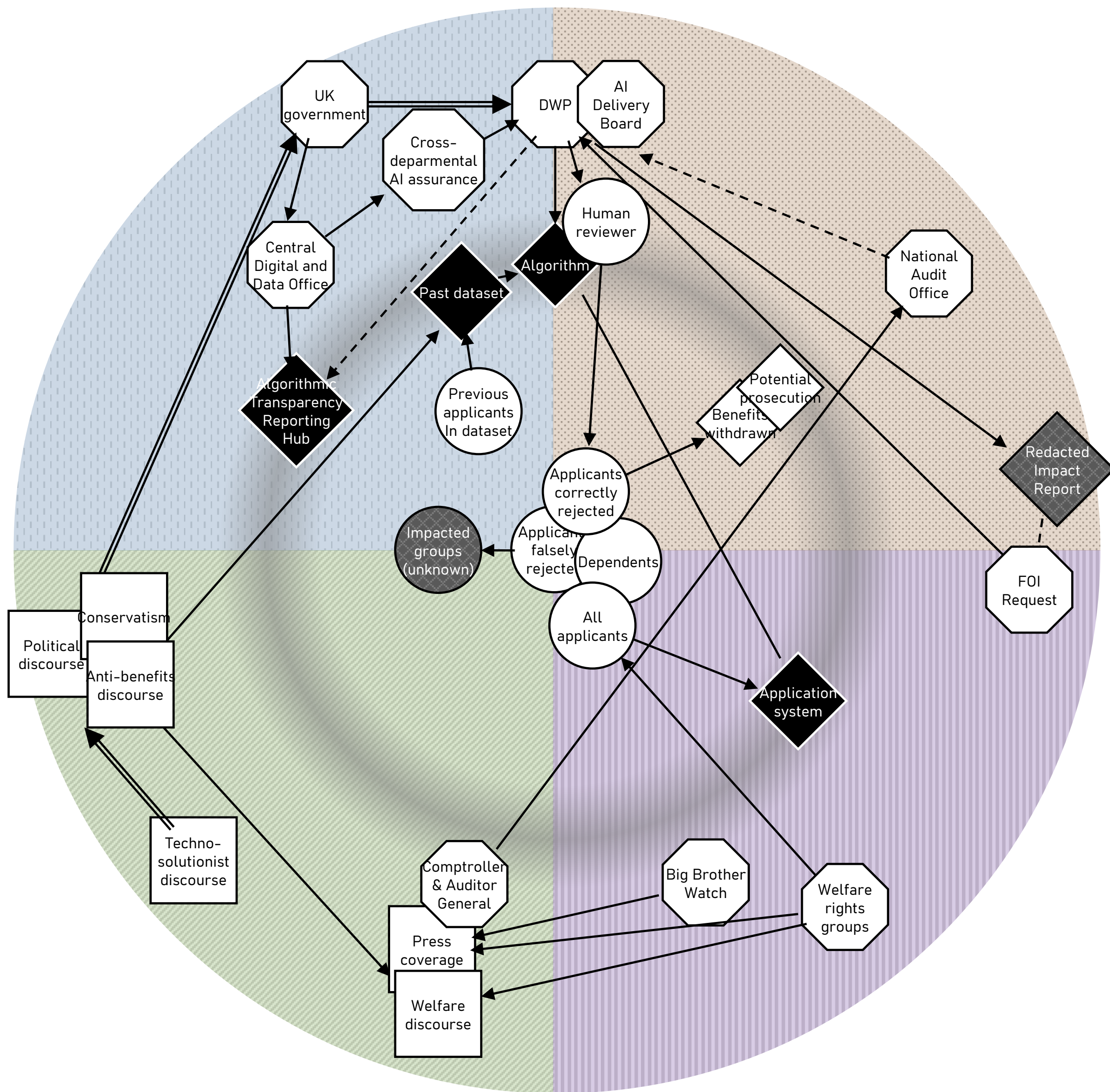

Figure 3: Map of the fraud risk assessment algorithm currently in use by the DWP, situated against transparency mechanisms

from previous blocks on the use of live facial recognition by the Scottish Parliament Justice Sub-Committee on Policing, which was influenced by academic and civil society groups calling for a moratorium, the debate has been reopened. The map shown in Fig.4 draws on information from the tripartite Working Group between Police Scotland, Scottish Police Authority and Scottish Biometrics Commissioner (Authority 2024), as well as consultations from previous Parliamentary discussions (on Policing 2020). These debates become situated in issues of Peelian principles (policing by consent), rights (minimum legal right rights compared to political discourse), political pressure (within Scotland and externally), different types of expertise (police, policy, academic and civil society, technology), normalisation and techno-solutionism (from other police forces or Police Scotland's existing use of retroactive facial recognition). Notable are statutory voids around the regulation of facial recognition, leaving major questions open to debate and demonstrating the potential influence of these different actors and discourses. The map demonstrates the social context of technology development, the different political, business, social and discursive interests that shape decision-making, and the relative power relations and domains of power in which different actors operate. Key highlights of this map are the potential points of intervention, given its development status and the open possibility of refusal. This is largely focused on the structural domain, and shifting the balance between the different elements of power. For example, currently Po-

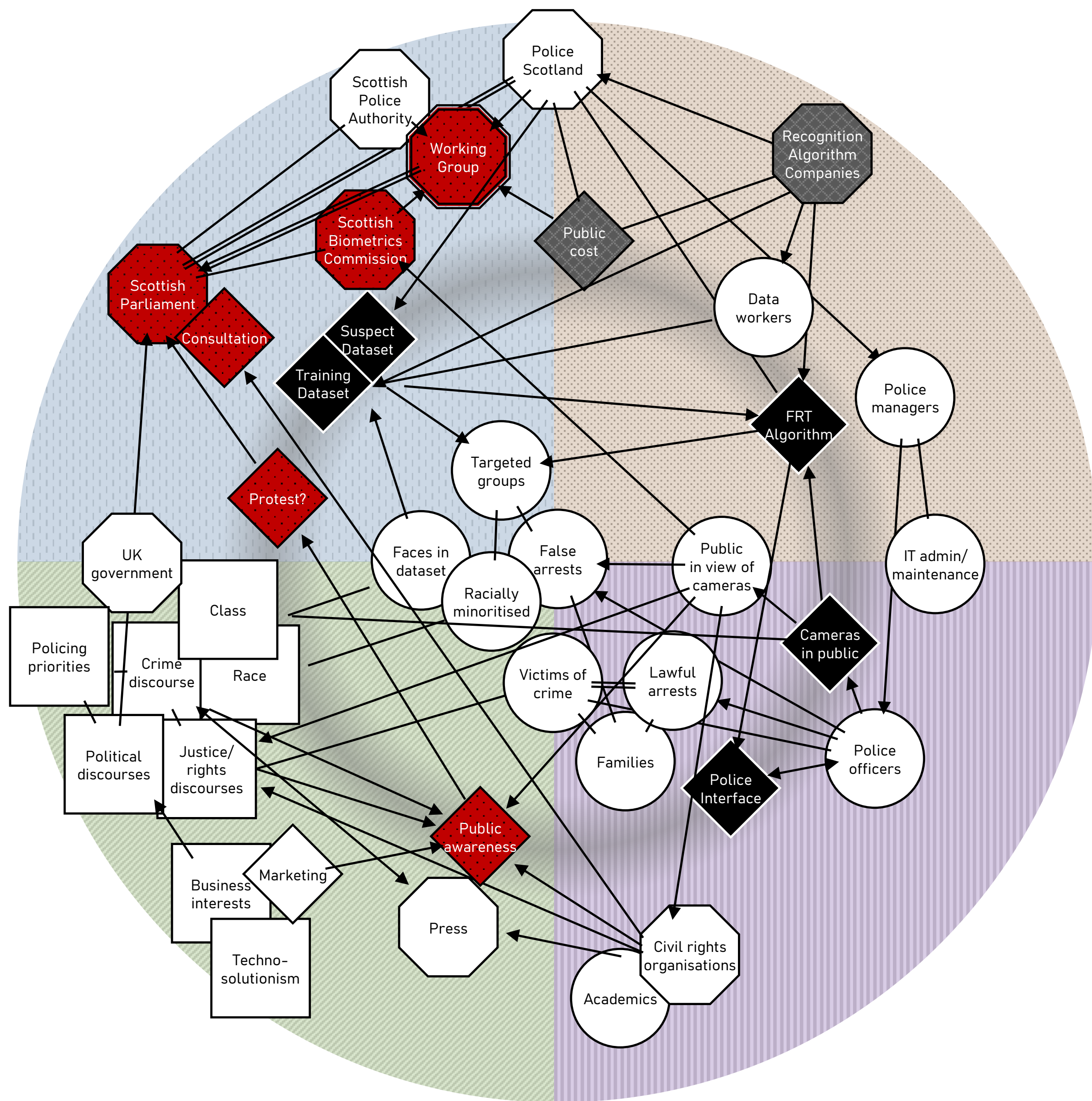

Figure 4: Map of decision-making and priorities feeding into proposed use of FRT by Police Scotland

lice Scotland is defining much of the narrative, supporting development of inequitable technologies with arguments for efficiency. The other parts of the working group could be pushed further into action, as the Scottish Parliament has done in the past through its committees and the biometrics commission in terms of enforcing existing rights. The purpose of the map is shown in the double-weighted lines highlighting the key tension points between police and political settings, and the tripartite working group as a key location of framing the debate moving forwards. Points where influence and pressure could be exerted are shown, as well as some of the unknown aspects around who would benefit (namely, the companies developing the technologies and the public funds they would be receiving). While the structural domain may appear the focus for action, the disciplinary domain should also be the focus for critique, with these corporate actors holding potentially more power if the systems gain political momentum towards approval and deployment. Directionality is particularly important in this map. For example, whereas a technical systems approach would show face data collected from the public in view of the camera, this mapping approach highlights how the cameras in public places are impacting on the public: an asymmetric power relation that undermines individual and collective rights by changing the nature of policing and places the public in a perpetual lineup, treating everyone as a potential suspect. These publics are also notably placed on the border of everyday and disciplinary domains of power, linking in further

connections of theory with, for example, biopolitics. This in turn completes a complex circling back to justice and rights discourses in the shifts implied by the use of the technologies. The map offers a critical view of the decision-making landscape in order to support researchers, policy-makers and activists in identifying gaps to be filled with evidence before any firmer decisions are made, before legislation is enshrined or technologies deployed and normalised. Its purpose is on the one hand to guide the decision-making and influencing process and on the other hand to pave the way for collective action should unjust uses be pushed forward by those in power.

Discussion

The mapping method presented has uses as an analytical process for historical or current AI systems, and as a way of thinking speculatively about AI systems in development or being proposed. The method can be applied to, among others: specific types of technology, such as facial recognition; landscapes of a field or subfield, such as computer vision; particular implementations or use cases, such as education, health, or public sector; labour, diversity, equity and inclusion (DEI) and influence within a particular industry or organisation; regulation and policy development including processes of consultation and influence. Across these different cases, the method allows us to ask critical questions about how a given AI system is designed, how it is deployed, who decides, and how discourses shape these decisions, all while focusing on how power affects the most marginalised.

We are not concerned with *tracing* technical systems or flows of data, but with *mapping* power relations in ways that are "open and connectable" (Deleuze and Guattari 2004). Our approach combines a theoretically driven framing based in the domains of power, and as such takes power relations as a starting point rather than an effect. The most affected, the most marginalised, are visually centred while the outer ring is discourses, thus unprivileging those with most direct decision-making power in the current forms of the technologies in question. It also embraces the need to decenter technical objects and their material components when discussing marginalisation (Peña Gangadharan and Niklas 2019). In doing so, we hope to enable a non-hierarchical approach that allows us to address power asymmetries and focus on approaches to change.

The method may be useful for researchers, policy-makers and regulators, educators, activists, journalists and civil society organisations, technologists, and even businesses for impact assessments. For example, for educators it is a way to teach students about power and social aspects of AI beyond technical systems, for researchers it is a useful step in visualising sociotechnical systems for analysis and presenting such analysis of power flows, for civil society organisations it presents a way to engage participants and also strategise specific places to target interventions, while for policy-makers it presents a useful visualisation method to highlight the need for cross-departmental cooperation and identify different groups to engage in consultation. The method is not a closed tool or formulaic checklist, but rather an open process for critical reflection and analysis. It should be undertaken purposefully in order to gain those critical insights for analysis and action.

There are, inevitably, limitations to the method as presented here. It may not be suitable in all circumstances. There may be systems for which details are so obscured that the map is more holes than points, and while identifying gaps in knowledge can aid in quests for transparency, there needs to be identifiable places in which to look for such information or mechanisms through which it may be found. Similarly, overfilling the map can limit its usefulness: the level of detail required for a comprehensive vision of a given AI system could potentially be too sprawling to be visually legible, with too many power relations to be able to tell what is going on.

The process of deciding where to collapse detail and where to elaborate may, however, become an integral part of constructing a given map, shaping the knowledge of the system in a situated way. The positionality and objectives of the mapper(s) - whether individual sets of researchers or those facilitating the map part of a wider engagement practice - is important. Just as the mapping method is a tool for data storytelling (Feigenbaum and Alamalhodaei 2020), so too is the process of mapping a weaving of a particular story from a particular perspective. Keeping the purpose in mind as a guide for where to condense and where to expand the detail will enable a more targeted and directed process. This should be combined with a focus on emphasising rather than concealing power dynamics, and ensuring those most affected remain visibile. The principles underpinning the map are opposed to universalising and totalising conceptions of AI, but embracing the openness, incompleteness and complex nature of such social relations may not be applicable in all situations. The aim, however, is to provide a useful way for different groups to be able to visually explore what an AI system is in social terms, to highlight tensions and power relations, and to identify points where changes (or refusals) can be made.

Conclusion

We have presented a novel method for mapping AI as a social assemblage of power relations that centres those most affected by a given system. We have offered a detailed discussion of the theoretical underpinning and critically reflexive development of the method, outlined the process by which we designed the visual presentation of the map, provided ways of populating the map, and given three examples of past (OfQual exams algorithm), present (DWP risk and fraud assessment algorithm) and future (Police Scotland's proposed use of live facial recognition technologies) implementations of the mapping method. The extended version provides larger images and a blank template for wider use. We hope that this will be useful to researchers and practitioners in a range of fields, as a way of making visible the complex power relations that are entrenched within AI development and deployment, analysing those power relations in the context of dominant structures and discourses, and identifying critical points of potential action towards more socially just assemblages of AI.

References

Ahmed, S. 2020. *Queer phenomenology: Orientations, objects, others*. Durham, NC: Duke University Press.

Aleman, E.; and Martinez, R. 2024. "It's Like They Are Using Our Data Against Us." Counter-Cartographies of AI Literacy. *Reading Research Quarterly*, 59(4): 665–677.

Ash, J.; Kitchin, R.; and Leszczynski, A. 2019. Introducing digital geographies. In *Digital geographies*, 1–10. London: SAGE Publications Ltd.

Authority, S. P. 2024. Live Facial Recognition. SPA Policing Performance Commitee Papers.

Baumann, Y.; Frenzel, J.; Guay, E.; King, L. I.; Megelas, A.; Renzi, A.; Rone, J.; Shahamati, S.; Vaughan, H.; Vukov, T.; et al. 2024. Data and Technological Spatial Politics. In *Dialogues in Data Power*, 159–185. Bristol: Bristol University Press.

BBC. 2020. A-levels: Unfair results claim after grades lowered.

Benjamin, G. 2022. # FuckTheAlgorithm: algorithmic imaginaries and political resistance. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 46–57.

Blunt, K. 2024. Nuclear-Powered AI: Big Tech's Bold Solution or a Pipedream? Amazon, Google and Microsoft are investing billions in nuclear power, but the projects are years away and rely on unproven technology. *The Wall Street journal*.

Booth, R. 2023. AI use widened to assess universal credit applications and tackle fraud. The Guardian.

Carraro, V. 2023. The C-word: how critical cartography, critical GIS and critical data studies can repoliticise disaster-related maps. *Disaster Prevention and Management: An International Journal*, 32(6): 1–13.

Clarke, A. E. 2003. Situational analyses: Grounded theory mapping after the postmodern turn. *Symbolic interaction*, 26(4): 553–576.

Clifton-Sprigg, J.; James, J.; and Vujić, S. 2020. Freedom of Information (FOI) as a data collection tool for social scientists. *PloS one*, 15(2): e0228392.

Collins, P. H. 2002. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. Abingdon: Routledge.

Crawford, K. 2021. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.

Crawford, K.; and Joler, V. 2018. Anatomy of an AI System.

Crenshaw, K. 1991. Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review*, 43(6): 1241–1300.

Davis, J. L.; Williams, A.; and Yang, M. W. 2021. Algorithmic reparation. *Big Data & Society*, 8(2): 20539517211044808.

Deleuze, G.; and Guattari, F. 2004. *A Thousand Plateaus: Capitalism and Schizophrenia*. London: Continuum.

Dieter, M.; Gerlitz, C.; Helmond, A.; Tkacz, N.; van Der Vlist, F. N.; and Weltevrede, E. 2019. Multi-situated app studies: Methods and propositions. *Social Media+ Society*, 5(2): 2056305119846486.

D'Ignazio, C.; and Klein, L. 2020. *Data Feminism*. Cambridge, MA: MIT Press.

Doyle, C.; Alvarez-Garcia, M.; Tracey, P.; Grill, G.; Whitney, C.; and Chambers, L. M. 2024. Reparations of the horse? Algorithmic reparation and overspecialized remedies. *Big Data & Society*, 11(3): 20539517241270670.

Emejulu, A.; and Sobande, F. 2019. *To exist is to resist: Black feminism in Europe*. Pluto Press.

Escobar, A. 2018. *Designs for the pluriverse: Radical interdependence, autonomy, and the making of worlds*. Durham, NC: Duke University Press.

Feigenbaum, A.; and Alamalhodaei, A. 2020. *The data storytelling workbook*. Abindgon: Routledge.

Fiorella, G. 2019. POSTSCRIPT: ACCESS IN THE ABSENCE OF FOI. *Freedom of Information and Social Science Research Design*.

Fricker, M. 2007. *Epistemic injustice: Power and the ethics of knowing*. Oxford: Oxford University Press.

Ganesh, M. I.; and Moss, E. 2022. Resistance and refusal to algorithmic harms: Varieties of 'knowledge projects'. *Media International Australia*, 183(1): 90–106.

Haraway, D. 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14(3): 575–599.

Harmon, K. A. 2004. *You Are Here: Personal Geographies and Other Maps of the Imagination (Imagined Maps Around the World, Collection of Artists Maps)*. Princeton, NJ: Princeton Architectural Press.

Heemsbergen, L.; Treré, E.; and Pereira, G. 2022. Introduction to algorithmic antagonisms: Resistance, reconfiguration, and renaissance for computational life. *Media International Australia*, 183(1): 3–15.

Hoffmann, A. L. 2019. Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society*, 22(7): 900–915.

hooks, b. 1984. *Feminist theory: from margin to centre*. Boston: South End Press.

Hung, K.-H. 2024. Artificial intelligence as planetary assemblages of coloniality: The new power architecture driving a tiered global data economy. *Big Data & Society*, 11(4): 20539517241289443.

Jones, E. 2024. FoI Request: AI Strategy Information. What do they know?

Kannabiran, K. 2006. A cartography of resistance: the National Federation of Dalit Women. *The Situated Politics of Belonging, London: Sage*, 54–74.

Kelly, A. 2021. A tale of two algorithms: The appeal and repeal of calculated grades systems in England and Ireland in 2020. *British Educational Research Journal*, 47(3): 725–741.

Kinkaid, E. 2020. Can assemblage think difference? A feminist critique of assemblage geographies. *Progress in Human Geography*, 44(3): 457–472.

Kitchin, R. 2014. *The data revolution: Big data, open data, data infrastructures and their consequences*. London: Sage.

Kobayashi, A. 2014. The dialectic of race and the discipline of geography. *Annals of the Association of American Geographers*, 104(6): 1101–1115.

Kobayashi, A. 2019. Issues of "race" and early radical geography: our invisible proponents. *Spatial histories of radical geography: North America and beyond*, 37–58.

Lehuedé, S. 2022. Territories of data: ontological divergences in the growth of data infrastructure. *Tapuya: Latin American Science, Technology and Society*, 5(1): 2035936.

Lin, W. 2015. Tracing the Map in the Age of Web 2.0. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 50(1): 41–44.

McKittrick, K. 2006. *Demonic Grounds: Black Women and the Cartographies of Struggle*. Minneapolis: University of Minnesota Press.

McKittrick, K. 2020. *Dear science and other stories*. Durham, NC: Duke University Press.

Milan, S.; and Treré, E. 2019. Big data from the South (s): Beyond data universalism. *Television & New Media*, 20(4): 319–335.

Milan, S.; and Treré, E. 2021. *Big Data from the South (s). An Analytical Matrix to Investigate Data at the Margins*, 1–21. Oxford University Press.

Milan, S.; Treré, E.; and Masiero, S. 2021. *COVID-19 from the Margins. Pandemic invisibilities, policies and resistance in the datafied society*, volume 40. Institute of Network Cultures.

Mollett, S.; and Faria, C. 2018. The spatialities of intersectional thinking: fashioning feminist geographic futures. *Gender, Place & Culture*, 25(4): 565–577.

Niang, S. M. 2024. in defence of what's there: notes on scavenging as methodology. *Feminist Review*, 136(1): 52–66.

Noble, S. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

of Work, D.; and Pensions. 2023. Annual Report and Accounts 2022-23 for the year ended 31 March 2023. UK Government.

OfQual. 2020. Awarding GCSE, AS I& A levels in summer 2020: interim report. UK Government.

Okidegbe, N. 2022. Of Afrofuturism, Of Algorithms. *Critical Analysis L.*, 9: 35.

Okidegbe, N. 2024. *Revisioning Algorithms as a Black Feminist Project*. University of California Press.

on Policing, J. S.-C. 2020. Facial recognition: how policing in Scotland makes use of this technology. Scottish Parliament.

Pasek, A.; Lin, C. K.; Cooper, Z. G. T.; and Kinder, J. B. 2023. *Digital Energetics*. Minneapolis: U of Minnesota Press.

Peña Gangadharan, S.; and Niklas, J. 2019. Decentering technology in discourse on discrimination. *Information, Communication & Society*, 22(7): 882–899.

Penn, J. 2021. Algorithmic silence: a call to decomputerize. *Journal of Social Computing*, 2(4): 337–356.

Pop Stefanija, A.; and Pierson, J. 2024. Repairing what's not broken–Algorithm repair manual as reflexivity device. *Convergence*, 13548565241261978.

Rakova, B.; Shelby, R.; and Ma, M. 2023. Terms-we-serve-with: Five dimensions for anticipating and repairing algorithmic harm. *Big Data & Society*, 10(2): 20539517231211553.

Rodríguez, C. 2017. Studying media at the margins: Learning from the field. In *Media activism in the digital age*, 49–61. Abingdon: Routledge.

Shaw, M. S.; Toliver, S.; and Tanksley, T. 2024. The internet doesn't exist in the sky: Literacy, AI, and the digital middle passage. *Reading Research Quarterly*, 59(4): 690–705.

Smith, G.; and Rustagi, I. 2021. When good algorithms go sexist: Why and how to advance AI gender equity. *Standford Social Innovation Review*.

Sovacool, B. K.; and Hess, D. J. 2017. Ordering theories: Typologies and conceptual frameworks for sociotechnical change. *Social studies of science*, 47(5): 703–750.

Suchman, L. 2002. Located accountabilities in technology production. *Scandinavian journal of information systems*, 14(2): 7.

Suchman, L. 2023. The uncontroversial 'thingness' of AI. *Big Data & Society*, 10(2): 20539517231206794.

Tanksley, T. 2023. Towards a Method of Black Feminist Archival Bricolage: Memory-Keeping within, beneath and beyond the Archive. *a/b: Auto/Biography Studies*, 38(2): 559–581.

Tanksley, T.; and Estrada, C. 2022. Toward a critical race RPP: How race, power and positionality inform research practice partnerships. *International journal of research & method in education*, 45(4): 397–409.

UNEP. 2024. A global foresight report on planetary health and human wellbeing.

van Angeren, J.; Alves, C.; and Jansen, S. 2016. Can we ask you to collaborate? Analyzing app developer relationships in commercial platform ecosystems. *Journal of Systems and Software*, 113: 430–445.

Walby, K.; and Luscombe, A. 2020. *Freedom of information and social science research design*. Routledge, Taylor & Francis Group.