6

Race, by Proxy

THAO PHAN AND SCOTT WARK

6.1 Introduction

In algorithmic culture, the problem of race is a problem of proxies. Though racism still continues in many familiar forms – through structural exclusion and oppression, through systems and policies that serve the neocolonial and/or white supremacist state, through words, actions and attitudes that demean and dehumanise – the tools and techniques that further these racist legacies and turn bodies into racialised subjects now operate in very different registers.[1] In part, this is an effect of what Frank Pasquale calls the 'black box society': a normalised culture of opacity in which large corporations shield their interests behind 'technical complexity, real secrecy, and trade secret laws'.[2] This culture of opacity makes it near impossible to know if and how racial (and other) discriminatory practices are taking place, whether intentionally – through, for example, the targeting or exclusion of ads for housing and education based on users' race;[3] or unintentionally – through, say, a differential pricing strategy that results in racialised and low-income groups consistently receiving higher prices for the same product.[4]

[1] For examples, see chapters in this volume by Aradau (Chapter 10), Tazzioli and Valdiva (Chapter 12), and Ardvisson, Gandorfer and McQuillian (Chapter 11).

[2] F Pasquale, 'The Black Box Society: The Secret Algorithms That Control Money and Information' in *The Black Box Society* (Harvard University Press 2015) 15.

[3] For example, Facebook's use of 'ethnic affinity' categories to allow advertisers to target and exclude content based on three broad categories: African American, US Hispanic and Asian American. For further discussion, see T Phan and S Wark, 'Racial Formations as Data Formations' (2021) 8(2) *Big Data & Society* 1; T Phan and S Wark, 'What Personalisation Can Do for You! Or: How to Do Racial Discrimination Without "Race"' (2021) 20 *Culture Machine* 1.

[4] For example, in 2012, a *Wall Street Journal* investigation found that the American retail conglomerate Staples adjusted online prices for users based on their proximity to competitor stores. The pricing mechanism took into account user postal codes, which led to users in low-income areas consistently receiving higher pricing. See J Valentino-DeVries, J

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

For some, this culture of opacity is an issue to be solved through better techniques of accountability and transparency. The aim is to render black boxes more explainable and observable; to make private interests and private algorithms more legible to public scrutiny.[5] Yet as many critical data scholars have argued, opacity isn't a straightforward condition that can be remedied through frameworks of 'shedding light' or 'opening up', but is in many cases an inherent aspect of the complex and fragmentary nature of algorithmic infrastructures themselves.[6] As political geographer Louise Amoore elegantly writes, 'the condition of giving an account that is never transparent or clear sighted is already the ethico-political condition with which we must live'.[7] That is to say, opacity defines the always partial, always situated, always mediated practices of representation and perception, algorithmic or otherwise.[8] In this context, knowing racism (and more importantly, resisting racism) means contending with these partial and opaque infrastructures, not necessarily as problems to be 'found and fixed' but as part of the many processes that mediate racialised life.[9]

Of interest for us here are the tensions that emerge when these opaque practices make use of ostensibly neutral data points to create clusters and associations that reconstitute categories of race in new and unusual ways. In previous work, we have claimed that systems driven by artificial intelligence (AI) are changing the nature of race as a category and racialisation as a process.[10] Alongside scholars such as

Singer-Vine and A Soltani, 'Websites Vary Prices, Deals Based on Users' Information' (*The Wall Street Journal*, 24 December 2012) available online at <https://bit.ly/3NX5MAM>; F Tramer et al, 'Discovering Unwarranted Associations in Data-Driven Applications with the FairTest Testing Toolkit' (arXiv, 8 October 2015) available online at <https://bit.ly/4fa4uyo>.

[5] B Rieder and J Hofmann, 'Towards Platform Observability' (2020) 9(4) *Internet Policy Review*.

[6] J Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3(1) *Big Data & Society* 1; P Dourish, 'Algorithms and Their Others: Algorithmic Culture in Context' (2016) 3(2) *Big Data & Society* 1; A Mackenzie, *Machine Learners: Archaeology of a Data Practice* (MIT Press 2017); A Mackenzie, 'From API to AI: Platforms and Their Opacities' (2018) *Information, Communication & Society* 1.

[7] L Amoore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others* (Duke University Press 2020) 19.

[8] Ibid 15–17.

[9] M Petersmann and D Van Den Meerssche, 'On Phantom Publics, Clusters, and Collectives: Be(com)ing Subject in Algorithmic Times (2024) 39 *AI & Society* 107.

[10] Phan and Wark, *supra* note 3.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

Marion Fourcade and Kieran Healy,[11] Oscar Gandy Jr,[12] Wendy Chun,[13] Louise Amoore[14] and Marie Petersmann and Dimitri Van Den Meerssche,[15] we have been interested in how modes of social sorting are today mediated through new techniques of supervised and unsupervised machine learning (ML). Whether placing people within market driven 'classification situations',[16] assessing and reassessing them along the unstable lines of racialised 'deep border' politics,[17] or differentiating them using the segregationist logic embedded in network science principles like 'homophily',[18] AI and ML systems inaugurate new ways of seeing and sorting bodies that has implications for understanding what a category like race is, how it functions, and how it is felt.

We can think of this as a post–big-data politics of race: what we have elsewhere called 'racial formations as data formations'.[19] Our focus has been to bring an attentiveness to the technicity of race's production to the study of race and algorithmic culture. This means grasping the specific apparatuses and the specific material modes through which race is mediated and produced. In this chapter, we build on this previous work by focusing on a central though rarely analysed technique: the use of proxies. In data science, a proxy is a feature that is used 'in place of a variable that cannot be measured'.[20] Proxies act as substitutes that can be used to infer, and therefore to act upon, an unknown piece of data. A proxy is thus a technique for measuring that which is in excess of perception. It reaches across the gap between the social and the computational not to make sense of the world, but to make it actionable within regimes of recognition, prediction and classification. Taken individually,

[11] M Fourcade and K Healy, 'Classification Situations: Life-Chances in the Neoliberal Era' (2013) 38(8) *Accounting, Organizations and Society* 559; M Fourcade and K Healy, 'Seeing Like a Market' (2017) 15(1) *Socio-Economic Review* 9.

[12] O Gandy Jr, *The Panoptic Sort: A Political Economy of Personal Information* (2nd ed., OUP 2021).

[13] WHK Chun, 'Queerying Homophily' in C Apprich et al (eds), *Pattern Discrimination* (1st ed., University of Minnesota Press 2019); WHK Chun, *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition* (MIT Press 2021).

[14] L Amoore, 'The Deep Border' (2021) 109 *Political Geography* 102547; L Amoore, 'Machine Learning Political Orders' (2022) 49(1) *Review of International Studies* 1.

[15] Petersmann and Van Den Meerssche, *supra* note 9.

[16] Fourcade and Healy (2017), *supra* note 11.

[17] Amoore (2021), *supra* note 14.

[18] Chun (2019), *supra* note 13.

[19] Phan and Wark, 'Racial Formations as Data Formations', *supra* note 3.

[20] G Upton and I Cook, 'Proxy Variable' in *A Dictionary of Statistics* (OUP 2014).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

these substitutions might not amount to much. Cascaded together, scaled up, and applied in concrete situations using AI and ML techniques, though, this use of proxies constitutes a *logic* by which the world can be *approximated* or modelled.

In what follows, we draw on a range of material from disparate fields, including critical race studies, media studies, science and technology studies, critical data studies, and the technical and legal literature on large-scale data processing to argue that this proxy logic provides us with a framework for understanding the situation we find ourselves in. This is a situation in which tools ostensibly designed to 'solve' the injustices caused by human racial bias have instead resulted in their recapitulation, and at worst, in new forms of discrimination. In this paradoxical scenario, attempts to remove race as an actionable data point results only in its entrenchment in a novel form. While algorithmic infrastructures promise to do away with crude categories like race in favour of more precise and emergent forms of clustering, race's differentiating logic nevertheless continues to shape experience.[21] We argue that proxy discrimination emerges as a consistent by-product of big data processing techniques, generating novel forms of 'colourblind' racism.

We then turn to the political implications of the proliferation of proxy logics. As scholars working with concepts like 'algorithmic governance'[22] or 'governance by data'[23] have noted, the increasing delegation of decision-making to AI and other automated systems effectively shields activity that is fundamentally political from scrutiny.[24] Here, the focus is often on the problem of opacity – what causes it, how to resolve it, and the risks it poses for democratic ways of life. But for us, it's also illuminating to see algorithmic governance as a mode of governance that operates *by proxy*.[25] The decisions we delegate to AI and ML systems are rendered

[21] Phan and Wark, 'Personalisation', *supra* note 3. Indeed, critical race scholars describe race as the 'master trope' of difference. See, for example, AG Weheliya, *Habeas Viscus: Racializing Assemblages, Biopolitics, and Black Feminist Theories of the Human* (Duke University Press 2014); HL Gates Jr, 'Editor's Introduction: Writing "Race" and the Difference It Makes' (1985) 12 *Critical Inquiry* 1.

[22] A Rouvroy and T Burns, 'Algorithmic Governmentality and Prospects of Emancipation' (2013) 177 *Réseaux* 163.

[23] F Johns, 'Governance by Data' (2021) 17(1) *Annual Review of Law and Social Science* 53.

[24] PD König, 'Dissecting the Algorithmic Leviathan: On the Socio-Political Anatomy of Algorithmic Governance' (2020) 33 *Philosophy & Technology* 476; S Issar and A Aneesh, 'What Is Algorithmic Governance?' (2022) 16(1) *Sociology Compass* 1. See also Goldenfein in this volume (Chapter 4).

[25] R Amaro, *The Black Technical Object: On Machine Learning and the Aspiration of Black Being* (Sternberg Press 2023).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

opaque not only by the sheer complexity of their underlying infrastructures, the overwhelming volumes of data that they process, or their arrogation of decision-making responsibilities by machines, but, more fundamentally, because they rely on proxy logics.

These logics embed another kind of political transformation: they arguably change the paradigm in which contemporary racial politics operates. Proxy logics use implicit forms of feedback and complex statistical transformations on our data to sort us into contingent groupings – what are sometimes called 'clusters'[26] or 'neighbourhoods'.[27] That is, one need only to have expressed the wrong combination of preferences or taken the wrong series of actions or have originated from the wrong place to be situated within the wrong kind of algorithmic space. In these instances, race is enacted not through modes of identification grounded in the reading of phenotype or genotype, but by latent associations between features of data.

Contemporary racial politics is associated, fairly or not, with what's often disparagingly called identity politics. On a personal level, each of us is committed to a political programme encompassed by this term. But we also want to argue that the abstract statistical vector spaces of AI and ML systems operate in an entirely incommensurable racial-political paradigm. To cluster or sort individuals into neighbourhoods, such systems don't discern our 'identity'. Rather, they draw inferences and approximations, creating clusters that group people together not on 'common characteristics, or meeting specified criteria' but instead on 'a spatialised proximity or distance'.[28] What becomes of race within this new paradigm? A paradigm which is defined not by a politics of identification and identity, but on a politics of approximation and proximity – in other words, when race is constituted by proxy?

6.2 On Proxies

A proxy is a substitute or a surrogate, a representative or a stand-in. The term proxy is used to describe a host of relations, from authorial power between different groups, actors and entities (e.g. delegates, legal intermediaries, members of a representative democracy), to forms of military

[26] Amoore (2021), *supra* note 14.

[27] Chun (2019), *supra* note 13. For further discussion on the effects of algorithmically mediated contingent grouping and their subversion of liberal subjectivity, see Petersmann and Van Den Meerssche, *supra* note 9.

[28] Amoore (2021), *supra* note 14, 6.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

and political power that bolster imperial statecraft (e.g. proxy wars and proxy forces), to the technical roles between different components in an infrastructural system (e.g. proxy servers, standardised measurements).[29] In its most basic form, a proxy signifies the act of delegation. It allows something to be treated *as if* it were another thing.[30] As media historian Dylan Mulvin has incisively argued, this seemingly simple process actually constitutes a key function of modern knowledge production. Identifying case studies and exemplars, working with samples, using testbeds and simulations: these methods all rely on proxy logics that create 'accessible representation' of otherwise chaotic or unknowable entities or events, which can then be acted upon *as if* they were the real thing.[31] 'This is the promise of proxies', writes Mulvin, 'the promise of creating controllable renditions of an unpredictable and unknowable world'.[32]

It shouldn't come as a surprise then that proxies feature prominently in AI and ML systems. These are systems that use big data-driven methods to calculate probabilities, model predictions, and ultimately, to tame uncertainty.[33] Here, proxies make the act of algorithmic prediction possible. Based on statistical methods, AI and ML attempt to solve problems through a logic of optimisation, a kind of rationality, to use Dan McQuillan's phrasing, that 'asserts that condensing [a context] into a calculation provides a superior kind of solution'.[34] Yet this rationality engenders a different kind of problem. As McQuillan pithily puts it, '[v]ery often we can't directly measure the things that matter the most, like happiness or health'.[35] To be able to render these kinds of phenomena amenable to calculation, 'algorithms have to rely on proxies'.[36]

Take, for example, recommender systems. The aim of a recommender system is to present people with things they want, to predict (or arguably,

[29] For a detailed discussion on the social and cultural functions of proxies, see D Mulvin, *Proxies: The Cultural Work of Standing In* (MIT Press 2021).
[30] See also Amaro, *supra* note 25 for an original and thought-provoking discussion on the strategic deployment of the 'as if' logic in relation to computation and the representation of black life.
[31] Mulvin, *supra* note 23, 4.
[32] Ibid.
[33] T Dryer, 'Algorithms under the Reign of Probability' (2018) 40(1) *IEEE Annals of the History of Computing* 93.
[34] D McQuillan, *Resisting AI* (Bristol University Press 2022) 24.
[35] Ibid 37–38.
[36] Ibid 38.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

to manufacture) desired behaviour.[37] Amazon recommends products you might like to buy. Netflix recommends movies you might like to watch. And Twitter recommends people you might like to follow. But these systems cannot produce a complete understanding of your preferences or desires. Actually determining what people want is much more complicated than it seems because people rarely have a coherent, rational or even consistent sense of what they want. Desire is variable and fluctuating, liable to be influenced by others and to change with context. It can contradict our professed values or be harmful to us, or can sometimes be mysterious or inexplicable – often, we simply don't know what we want. Desire, as it turns out, is elusive and immeasurable. So, in the face of the impossibility of knowing desire, recommender systems instead make a substitution based on other signals – usually behaviours that can be observed and measured in computational terms.

For most platforms, this takes the form of a wide range of explicit and implicit signals. These range from types of explicit engagement – such as clicks, views, shares, reactions or 'dwell time' – through to other kinds of implicit feedback – such as inferences drawn from people in your network. In data science, this is called determining 'revealed preference', and is premised on the assumption that observable actions can be used to reconstruct the motivation and preferences of a person. This approach is informed by now outdated nineteenth and twentieth century economic theories of consumer behaviour, specifically the characterisation of the consumer as a rational actor whose actions are driven by frameworks like informed decision-making and a utilitarian agenda.[38]

Determining revealed preference operates through a series of proxy substitutions: clicks as a proxy for engagement; engagement as a proxy for rational behaviour; rational behaviour as a proxy for desire; desire as proxy for identity and the self. While desire might not be amenable to calculation and to optimisation, its proxies are. These proxies allow otherwise mutable and amorphous phenomena to be the subject of calculative techniques. In the words of media scholar Wendy Chun, proxies allow us to 'touch the unknown: they extend the unknowable, by capturing or "synching up with" what is not there'.[39] Proxies are thus

[37] M Andrejevic, *Automated Media* (Routledge 2019); N Seaver, *Computing Taste: Algorithms and the Makers of Music Recommendation* (University of Chicago Press 2022).

[38] L Thorburn, J Stray and P Bengani, 'What Does It Mean to Give Someone What They Want? The Nature of Preferences in Recommender Systems' (*Understanding Recommenders*, 11 May 2022) available online at <https://bit.ly/48Ayr8m>.

[39] Chun (2021), *supra* note 13, 137.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

the inevitable effect of tools that seek prediction despite unknowability, that manufacture order amidst chaos. They substitute *what we want* to measure with *what we can* measure, allowing us to act upon data about the world *as if* it represented the world itself.

On one level, we could argue that all data are proxies. They are representations (often numerical values) that are taken as stand-ins for certain phenomena in the world. Making knowledge with data is, then, always already a process of making substitutions. With AI and ML systems, the vast array of applications, the scale of tasks, the complexity of processes, and the increasing stakes of outcomes now assigned to data-driven systems have new implications for lives caught within these regimes. To understand the harms and risks of 'global governance by data', then, one must also engage with the opacities and complexities inherent to proxies.

6.3 Proxy Discrimination As Colourblind Racism

One of the cruel ironies of the data-driven age is that systems designed to solve or reduce harms, like racism and racial discrimination, have in practice entrenched or expanded their effects. A key example of this is proxy discrimination.

Direct discrimination on the basis of gender, race, sexuality or other protected attributes is usually prohibited by law. For instance, the use of race as an explicit variable in the determination of decisions regarding credit eligibility, employment, or housing. But discrimination can also occur indirectly through the use of correlations between protected classes and other attributes. For instance, the use of postcodes as a proxy for race, income as a proxy for gender. This is known as proxy discrimination, and while it has historically been understood as a kind of direct discrimination,[40] the widespread application of algorithmic and automated tools into a growing number of high-stakes decision-making processes has seen a new kind of proxy discrimination emerge.

Here, discrimination is a by-product of AI models that draw associations, inferences, and other statistical relationships between correlated sets of data.[41] For example, a 2012 Wall Street Journal investigation found that the American retailer Staples adjusted online prices for users

[40] AER Prince and D Schwarcz, 'Proxy Discrimination in the Age of Artificial Intelligence and Big Data' (2020) 105(3) *Iowa Law Review* 1257, 1262.

[41] S Barocas and AD Selbst, 'Big Data's Disparate Impact' (2016) 104(3) *California Law Review* 671, 677.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

based on their proximity to competitor stores. The intention was that customers living within twenty miles of a rival store, such as Office Depot, would receive greater discounts and deals. What was overlooked, however, is that access to infrastructure, such as retail stores, is highly correlated with geography and postal codes, which in turn are correlated with income and race. The result was that low-income and racialised neighbourhoods consistently received higher prices for Staples products.[42]

In another infamous example, Amazon introduced an AI-powered recruitment tool to identify promising employees based on their resumes. The tool was quickly retired, however, after the company found it systematically downgraded female applicants based on proxy indicators. Applicants from historically female colleges, who named certain hobbies, or who even used the word women (such as 'women's chess champion') in their resumes were given lower scores.[43] In both these instances, neither gender nor race were included as explicit variables, but the patterns discovered by the AI and ML tools nevertheless produced discrimination against women and racialised groups as a second order effect.[44] Even as race recedes, discrimination returns, in novel, algorithmic, gendered and/or racialised guises.

How do we address these pernicious forms of sexism and racism? On the one hand, we could argue, as many scholars have, that these racist and sexist outcomes are not simply the 'unintended consequences' of working with big data, but are the all too predictable result of systems designed to recognise existing patterns in social relationships.[45] They reveal what Solon Barocas and Andrew Selbst call 'the simple fact of inequality'.[46] As such, any efforts to 'fix' these systems by creating more precise models or capturing more data will merely replicate, or in many cases exacerbate, the prejudices and injustices that define the institutions governing life as we know it.[47]

[42] Valentino-DeVries, Singer-Vine and Soltani, *supra* note 4.

[43] J Dastin, 'Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women' (*Reuters*, 10 October 2018) available online at <https://bit.ly/3YPGfQq>.

[44] We would like to thank Julie Cohen for providing this incisive comment on an earlier draft of this work.

[45] Apprich et al (eds), *supra* note 13; Chun (2021), *supra* note 13; C O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown Publishing Group 2016).

[46] Barocas and Selbst, *supra* note 41, 691.

[47] R Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (1st ed., Polity 2019); O'Neil, *supra* note 45.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

On the other hand, we might take a broader look, acknowledging that these AI systems have themselves been introduced as a solution to the problem of good ol' fashioned human racism that pervades institutional decision-making. With this in mind, we can examine the almost naive logic behind this approach. Chun describes this succinctly as the assumption that 'racism and discrimination naturally stems from human recognition, and thus the cure must be the erasure of all visible markers of difference. If only we got rid of the markers of race – it is presumed – then all would be good'.[48]

Building on the work of sociologist of race Eduardo Bonillo-Silva, Chun argues that this approach constitutes a kind of 'colourblind racism' that over-invests in the neutralising force of technology. Like Chun, we also wish to interrogate this 'race-free equals racism-free' formula.[49] Where Chun looks historically at histories of eugenics and segregation and their role in shaping the principles of network science, we wish to stay with the question of race as it is mediated and constituted today. Specifically, we wish to understand how processes of racialisation and race-making are now contingent on proxies and logics of proxification, and the implications this has for racialised subjects who find themselves on the receiving end of these 'unintended' racist effects.

This is a significant conceptual move because problems like proxy discrimination have proven themselves to be incredibly vexed issues for computer scientists, policy makers, and critical scholars alike. Try as we might, we can't seem to stop proxy discrimination from happening, even when – or, perhaps, explicitly when – race is actively excluded as a feature or variable within a dataset. Despite the great equalising promise of these technological systems, racism seems to always find a way to manifest.

Indeed, even if we look at efforts to remediate proxy discrimination through technical procedures of verification and audits, we find proxies everywhere. For example, in a widely cited paper on 'Auditing Black-Box models for indirect influence', computer scientist Philip Adler[50] and his team developed a method to determine whether specific features in a dataset (like gender or race) influence the outcomes of a model's prediction, and if so to what degree. Their focus was specifically on proxy discrimination in black box models. To do this, they used two techniques:

[48] Chun (2021), *supra* note 13, 17.
[49] Ibid 22.
[50] P Adler et al, 'Auditing Black-Box Models for Indirect Influence' (2018) 54(1) *Knowledge and Information Systems* 95.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

(1) synthetic data, and (2) models of models. Both these techniques can be interpreted as themselves relying on proxy logics.

Synthetic data is an artificially manufactured dataset. It is data that is created algorithmically rather than derived from actual real-world measurement, and is often used as a stand-in for tests to validate or train ML models. It looks and has the attributes of a real-world dataset, but has the added bonus of being already cleaned as well as side stepping issues to do with data privacy. So, for example, synthetic data is often used in healthcare contexts because it allows for the public use of record-level data without having to worry about patient confidentiality because the features represented are not derived from real people but proxy subjects.[51]

Similarly, models of models are used as a stand-in when one is precluded from accessing a model's backend. They are specifically used in cases of black boxes, when a researcher is unable to tweak and rerun an algorithm to see what effect changing this or that variable will have. Instead, the approach is to create a model of a model: a kind of approximation of how you think the model works based on what is observable. For example, in their study examining the visual clustering techniques of the social media platform Instagram, Nic Carah, Daniel Angus and Jean Burgess created their own machine vision system to undertake an unsupervised classification task on over 13,000 images posted to the platform.[52] They then analysed the features of the clusters created in an effort to understand how images become operationalised within ecologies driven by personalisation and recommendation for online advertising. They described this technique as a kind of 'critical simulation' that allows them to perform 'a speculative analysis'.[53] Similarly, Adler et al used this technique to audit the now infamous Northpoint COMPAS software to try and examine exactly how race, as a feature, impacted on outcome scores for recidivism.[54] Their method examined the interaction between model and data, measuring the differences and sensitivities when certain features were obscured. This technique of auditing a proxy model as a stand-in for a black box system can then

[51] BN Jacobsen, 'Machine Learning and the Politics of Synthetic Data' (2023) 10(1) *Big Data & Society*; M Richardson, 'Witnessing Algorithms and the Paradox of Synthetic Media' (*Circus Bazaar*, 1 November 2022) available online at <https://bit.ly/48EozKE>.

[52] N Carah et al, 'Tuning Machines: An Approach to Exploring How Instagram's Machine Vision Operates On and Through Digital Media's Participatory Visual Cultures' (2022) 37(1) *Cultural Studies* 20.

[53] Ibid 23.

[54] Adler et al, *supra* note 50.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

be used within broader calls for investigations into potential indirect discrimination.

What this example illustrates is that even in our efforts to observe or measure proxies and their effects, we rely on techniques that are themselves premised on proxy logics. When it comes to knowing and being with algorithms, then, it really is proxies all the way down.

6.4 The Proxification of Race

In the slippage between absence and presence, between the erasure of race and the persistence of racist effects, we find the kernel of what we want to call a logic. This logic helps us to explain not only how racism continues to manifest within systems designed to remedy against prejudice, but to understand precisely how race is being acted upon and transformed in this process. We want to call this logic 'proxification'. Proxification operationalises the *as if*. It puts it to work in systems and structures that seek to render the world into computational terms to make it more amenable to order and control. While the logic of proxification is present in algorithmic systems that implement rule-based sorting mechanisms, it is supercharged by the spread of AI and ML techniques.

Of particular interest to us is the role of unsupervised ML in surfacing latent and unrecognised connections. Carah, Angus and Burgess provide an efficient explanation distinguishing between supervised and unsupervised ML systems, describing supervised ML as being 'oriented toward classification, whereas unsupervised approaches are oriented toward association'.[55] Writing specifically on machine vision systems, they state:

> Supervised approaches to machine vision involve training a machine to recognize pre-defined features, such as people, objects, logos, or text, inside an image. A supervised algorithm is told in advance what the 'features' are and is trained to correctly identify those features within a specified error tolerance . . . In contrast, unsupervised approaches involve designing a machine to inductively recognize recurring patterns from the latent properties of input data.[56]

Because they draw on data's 'latent properties', this inductive form of pattern recognition means that associations and connections can be generated in ways that are often both strikingly novel and extremely

[55] Carah et al, *supra* note 52, 26.
[56] Ibid 27.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

difficult to reverse-engineer. According to Adrian Mackenzie, these latent connections arise when the features that data exhibit are mapped on to volumes in 'vector space'.[57] In this vector space, data need not be grouped together on the basis of explicit shared characteristics. Data might contain so many dimensions – variables – that what they do or do not share might not be particularly clear. Rather than sorting data into groups based on unambiguous characteristics, then, a data subset's 'similarity and belonging' must instead be determined by 'measures of proximity or distance, on flat loci that run as vectors through the space'.[58]

As Louise Amoore notes, this kind of statistical transformation allows AI and ML techniques to apprehend emergent 'clusters' defined by multiple shared characteristics.[59] The use of clusters transforms the logic by which subjects are sorted into groups, because the problem frame is no longer centred strictly around identification but around alignment and association. She writes:

> When a new entity appears before the model, the primary question is not who or what are they (a daisy or a rose, a dog or a cat) but rather 'to which cluster are their attributes most closely aligned' . . . To belong to a cluster is not about resemblance, common characteristics, or meeting specified criteria, but is instead about spatialised proximity and distance.[60]

Clusters thus enact new forms of grouping. Objects are grouped based on their proximity to the centre of the cluster and classified depending on threshold values. This is how an individual might be assigned a risk score, or a number that indicates the degree to which they might have a propensity to default on a loan, defraud welfare services, commit a crime, or pose an existential threat to the state.[61]

Proxies are fundamental components of the measures of proximity that AI and ML systems use to draw inferences and make determinations. But while clusters might generate particular statistical predictions rendered as measures – of, for instance, probability, or of confidence in a probability measure – their outputs have to be understood as imperfect approximations. Here, the slippage between what we *want* to measure – a determinate state of affairs – and what we *can* measure – on the basis of the data we have available, the infrastructures we have established, the

[57] A Mackenzie, *Machine Learners: Archaeology of a Data Practice* (MIT Press 2017) 51.
[58] Ibid 73–74.
[59] Amoore (2021), *supra* note 14, 5.
[60] Ibid 6.
[61] Ibid; J Cheney-Lippod, '*Jus Algoritmi*: How the National Security Agency Remade Citizenship' (2016) 10 *International Journal of Communication* 1721.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

compute resources we have access to, and so on – is what transforms the proxy from a minor substitution into an ordering logic.

This logic emerges in the slippage between measurement and representation. While AI and ML systems purport to measure the world, what they actually offer is an increasingly granular statistical representation of that world. Amoore captures this slippage when she notes that neural networks 'hold[] out an alluring promise to the state; it too could decompose and represent everything, if only it could harness high-dimensional data to resolve the intractable difficulty that is politics'.[62] Amoore's wry observation recalls the age-old representational problem of resolution: a map would become the territory if it were rendered in full detail; politics, by analogy, would no longer be necessary if one's model was robust enough.

While we know that AI and ML systems can't re-render the world whole and complete, producing it identically, the slippage between their incapacity to do so and the confidence we place in them to make decisions that actively shape the world has effects of obfuscation. We use them to make not only single, important decisions. We also use them to make countless small, seemingly inconsequential decisions that, in the aggregate, have profound effects. In the process, we lose sight not only of how these effects might scale to produce emergent and unpredictable outcomes, but also how they introduce new forms of difference into the world at large.

McQuillan highlights this capacity when he describes AI as a 'technology of racialization' which only 'knows how to discriminate into output classes' by 'calculating distances, by determining some abstract metric of difference as distance'.[63] Distance differentiates. Distance carving undifferentiated wholes into groups. Distance is a measure of *dis*similarity. Distance, in other words, is able to operate not only as a converse measure to proximity, but *as a means of discriminating between* groups. Distance is, itself, able to operate as a statistical proxy – for, amongst other things, *race*. The kinds of discrimination, the feeling of these racist effects aren't necessarily new – but the technical means that they are produced are. To conclude, we want to connect these novel modes of racialisation to governance, examining how proxification situates race right at the heart of the transformations that governance itself has been undergoing.

[62] Amoore (2021), *supra* note 14, 3.
[63] McQuillan, *supra* note 34, 75.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

6.5 Conclusion: From a Politics of Identity to a Politics of Proximity

In his influential book on the impact of statistical reasoning on the emergence of modern statecraft, the historian of science Alain Desrosières argued that statistics provided governments with the capacity to produce an 'equivalence space' in which a state's resources – people, resources, armies, trade goods, capital – could be made commensurable and, therefore, calculable.[64] We can consider AI and ML as continuing this operational function, expanding and transforming this equivalence space into a vector space from which order is created and decision-making is derived.

These transformations have several implications for understanding contemporary governance. Numerous scholars have analysed how digitisation, the proliferation of algorithmic systems, the explosion of data collection techniques and datasets and the emergence of the platform as a key twenty-first-century infrastructure have transformed how governance is conducted by states and broadened who might be said to govern contemporary society.[65] These techniques are increasingly integrated into formal systems of governance, promising to either make its institutions more efficient or streamlined,[66] or else to mediate the distribution of increasingly scarce resources in increasingly stringent times.[67] They shape our shared sense of social reality[68] by introducing novel modes of sorting and ordering individuals into social collectives.[69] But if we understand governance to be a process of social ordering that is constrained by rules[70] that isn't solely carried out by states,[71] we can begin to see how seemingly innocuous techniques – such as the use of algorithms

[64] A Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning* (Harvard University Press 1998) 21.

[65] JE Cohen, 'The Biopolitical Public Domain' in *Between Truth and Power* (OUP 2019); Gandy Jr, *supra* note 12. See also chapters in this volume by Julie Cohen (Chapter 2) and Jennifer Cobbe (Chapter 8).

[66] M Veale and I Brass, 'Administration by Algorithm? Public Management Meets Public Sector Machine Learning' (SSRN 2019) available online at <https://bit.ly/40xMBVO>.

[67] S Day, C Lury and H Ward, 'Personalization: A New Political Arithmetic?' (2023) 24(2) *Distinktion: Journal of Social Theory* 167; McQuillan, *supra* note 34.

[68] N Just and M Latzer, 'Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet' (2017) 39(2) *Media, Culture & Society* 238.

[69] Fourcade and Healy (2017), *supra* note 11.

[70] C Katzenbach and L Ulbricht, 'Algorithmic Governance' (2019) 8(4) *Internet Policy Review* 1.

[71] Issar and Aneesh, *supra* note 24.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

to control the content we see online – fundamentally reshape what governance is and how it operates.

In their formative work on algorithmic governmentality, for instance, Antoinette Rouvroy and Thomas Berns have claimed that algorithms usher in a mode of governance that doesn't act on political subjects. Instead, by operating via statistical means, algorithms enable a mode of governance that attempts to 'act on', and thereby shape, the 'environment' that such subjects inhabit.[72] In this, it is supposed to constitute 'governance without a subject, but not without a target',[73] and can be linked to governance techniques derived from behavioural science, such as 'nudging'.[74] In a more recent intervention into this debate, Fleur Johns has also argued that governments' access to gigantic – and ever-increasing – pools of data is itself having a transformative effect on governance. Johns calls this alternative concept 'governance by data', arguing that this phrase 'denotes instead the propensity for the gathering, assemblage, formatting, analysis, transmission, and storage of digital data to have governance effects'.[75]

We are agnostic about ascribing these transformations in governance to algorithms or to data – or to another agent altogether, such as the platform.[76] These frameworks are all compelling; importantly, they aren't mutually exclusive. What we want to focus on is not what ultimately determines this transformation in governance, but what it means for how politics is done – and more specifically, how *racial* politics play out.

These systems subject politics itself to a logic of proxification. In his recent work on machine learning, Ramon Amaro makes an illuminating observation about AI and ML systems. AI and ML systems' power to discriminate between data, surface latent connections, identify clusters, and automate the production of outputs amounts to a form of proxy politics, because such systems are often implemented as a proxy *for* 'collective and individual decision making'.[77] By acting on data rather than subjects, this mode of governance can operate *as if* its effects were rational outcomes of objective processes.

[72] Rouvroy and Burns, *supra* note 22, 7.
[73] Ibid 10.
[74] K Yeung, '"Hypernudge": Big Data as a Mode of Regulation by Design' (2017) 20(1) *Information, Communication & Society* 118.
[75] Johns, *supra* note 23, 55.
[76] J Van Dijck, T Poell and M De Waal, *The Platform Society: Public Values in a Connective World* (OUP 2018); BH Bratton, *The Stack: On Software and Sovereignty* (MIT Press 2015).
[77] Amaro, *supra* note 25, 32.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

This is why Rouvroy and Berns argue that this emergent form of governance crystallises a 'certain type of (a)normative or (a)political rationality founded on the automated collection, aggregation, and analysis of big data so as to model, anticipate and pre-emptively affect possible behaviours'.[78] In general terms, this proposition points to how the implementation of algorithmic systems changes the way that governance operates. Rather than taking individuals – whether construed as citizens or users – as the direct *subjects* of its operations, Rouvroy and Berns argue that this mode of governance affects the relations between the data produced by subjects.[79]

By delegating a portion of our freedom not only to decide, but to *deliberate* over the justness of decisions to the AI and ML systems that constitute algorithmic governance, we are not just eliding the deeply political nature of what these systems do but encoding new forms of inequality. As Shiv Issar and Aneesh Aneesh argue, as these systems become embedded, they constitute a new 'societal infrastructure' that 'allow[s] inequalities to become durable'.[80] For some scholars and commentators, such inequalities might take the form of unequal access to resources, or emerge when unfair decisions are occluded by AI and ML's a-political and a-normative veneer. For others, they derive from data that represents already-existing, concrete inequalities, which are reproduced by AI and ML systems. This is bad enough, but our claim is that the stakes are even higher.

Earlier, we described proxies as bits of data that can be used in place of variables that cannot be measured. We also referred to these bits of data as 'minor substitutions' that, in and of themselves, don't necessarily produce outsized political effects. But as the examples we've outlined in the prior sections illustrate, such minor substitutions collectively conspire to perform a sleight of hand. The more AI and ML arrogate decision-making powers, the more they reconstitute political problems within paradigms of clustering, proximity, association, probability, and pre-emption, the more they risk embedding new apparatuses of domination and discrimination into our digitised societies.

What does this mean for contemporary racial politics? In a recent essay on the impact that deep learning is having on the governance of borders, Louise Amoore argues that 'the deep learning techniques of our

[78] Rouvroy and Burns, *supra* note 22, 10.
[79] Ibid 4.
[80] Issar and Aneesh, *supra* note 24, 8.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

contemporary moment are making race signify something different: a set of emergent proximities and distances generated through clustering'.[81] We not only agree with this statement, but think AI and ML extend its logic beyond the borders demarcated by the 'border'. Insofar as AI and ML are made of proxies all the way down, they remain open to the possibility of being corrupted by racialised discrimination from the inside out.

As we have noted, the populations calculated and constructed by AI and ML systems are premised not on discrete, shared characteristics – on 'identity' – but on measures of alignment determined using complex, automated machine learning techniques: on what's referred to as the 'proximity' between individuals,[82] which might be derived from a huge number of parameters including not only an individual's characteristics (whom they are like), but, increasingly, preferences inferred using behavioural signals.[83] It's in this novel mode of calculating persons that a new racial politics takes root: as proxies cascade, they reorder the relationship between cause and effect that has heretofore underpinned the modern construction of race. Where it was once the body that determined the pattern (genotype or phenotype that acted as the signal for racial identification and belonging), it is now the pattern that determines the body (patterns of behavioural data that are used to create clusters, associations, correlations with a label representing identity).

Beyond engendering shifts in how communities or polities are governed, algorithms change the politics of race because they change how we – the collectivity, understood as broadly as possible, of those subjected *to* racialisation – are seen and represented and how we can act or be acted upon. This is why the interplay of opacity and representation or perception that Amoore highlights is the necessary vanishing point for the analysis we want to present in this essay. What proxies – and the AI and ML systems which rely upon them – engender then is a shift from a politics of identity to a politics of proximity.

We think a fair amount of paranoia is warranted in this situation. Rather than dealing with a biased state institution or service, or a biased monolithic corporation, our contemporary situation confronts us with countless services, premised on data processing using techniques like AI and ML, that *may* subject us to incorrect racialisations with unfair,

[81] Amoore (2021), *supra* note 14, 7.
[82] Ibid; Mackenzie, *supra* note 57.
[83] Day, Lury and Ward, *supra* note 67.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008

discriminatory, or unequal outcomes. This influence is produced, enabled, and made durable, in novel form, by the very proxy logics on which these systems rely.

Using proxies allows engineers to declare their AI and ML systems 'colourblind', because they don't contain data on categories such as race. And yet racism persists in these systems, mutating – via proxies – to take on new and even more opaque forms. As proxies cascade, turning minor substitutions into a broader logic and normalising data-driven techniques of *discriminating between* different entities, they increase the risk that we submit ourselves to a whole new racial paradigm.

In this way, proxies should not be understood as a problem of algorithms, but rather the intractable condition that makes algorithmic life possible – and new, racialised, modes of data-driven discrimination inevitable. The key point we want to make in concluding is this: to resist this inevitable situation, we need to learn a new set of political techniques equal to a politics of proximity. This doesn't mean that identity politics is now redundant, or that its aims and goals should be discarded. The two are hardly mutually exclusive, and racial discrimination is so multiplicitous and varied that we need all the political strategies we can get. But it does mean that the tactics of identity politics aren't equal to the task of resisting new modes of racialisation and racial discrimination premised on logics of proxification.

This sleight of hand contained within the proxy is both an act of disavowal and erasure. To stay with race, we need to refuse its logic. The form that this refusal will take is, for us, perhaps the most urgent question we, that racialised collectivity, need to answer today.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 18:38:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.008