

Can *critical policy studies* outsmart AI? Research agenda on artificial intelligence technologies and public policy

Regine Paul

To cite this article: Regine Paul (2022): Can *critical policy studies* outsmart AI? Research agenda on artificial intelligence technologies and public policy, *Critical Policy Studies*, DOI: [10.1080/19460171.2022.2123018](https://doi.org/10.1080/19460171.2022.2123018)

To link to this article: <https://doi.org/10.1080/19460171.2022.2123018>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 19 Sep 2022.



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



Article views: 91



View related articles [↗](#)



View Crossmark data [↗](#)

Can *critical policy studies* outsmart AI? Research agenda on artificial intelligence technologies and public policy

Regine Paul

Department of Government, University of Bergen, Bergen, Norway

ABSTRACT

The insertion of artificial intelligence technologies (AITs) and data-driven automation in public policymaking should be a metaphorical wake-up call for critical policy analysts. Both its wide representation as techno-solutionist remedy in otherwise slow, inefficient, and biased public decision-making and its regulation as a matter of rational risk analysis are conceptually flawed and democratically problematic. To 'outsmart' AI, this article stimulates the articulation of a critical research agenda on AITs and public policy, outlining three interconnected lines of inquiry for future research: (1) interpretivist disclosure of the norms and values that shape perceptions and uses of AITs in public policy, (2) exploration of AITs in public policy as a contingent practice of complex human-machine interactions, and (3) emancipatory critique of how 'smart' governance projects and AIT regulation interact with (global) inequalities and power relations.



KEYWORDS

Artificial intelligence technologies; automated decision-making; interpretivism; practice; critical political economy; critique

Introduction

Artificial intelligence technology¹ (AIT) is arguably not just any new technology but a veritable *ecology* transforming 'our lives, our interactions, and our environments' (Cath et al. 2018, 4). As an environment that shapes ever more aspects of our social habitat – from our daily consumption and communication patterns to how we can access services and even the way we date – the AIT ecology also affects *public* decision-making: (1) public administrations have increased and deepened their uses of AITs to detect patterns, sort populations, and predict behavior² *alongside* (2) equally growing global efforts to regulate AITs.

- (1) Drawing on big data analytics³ the use of AITs to automate governmental decision-making partially or fully now includes chat bots at user interfaces with public services to manage large volumes of citizen contact, automated case work to detect social security or tax fraud, smart diagnostics and 'nudging' in health care, predictive policing, facial recognition, and biometric identification at the 'smart' border, or dialect recognition in the processing of asylum claims (Henman 2020; Jeandesboz 2021; Kaufmann, Egbert, and Leese 2019; Prainsack 2020). Such public

CONTACT Regine Paul  regine.paul@uib.no  Department of Government, Faculty of Social Sciences, University of Bergen, Postboks 7802, 5020 Bergen, Bergen, Norway

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

uses of AI often represent *technological solutionism*, where actors construct problems as urgent and solvable through specific technical devices (Morozov 2014). In a context of public sector austerity, New Public Management hypes, but also increasing policy complexity, techno solutionism around AITs makes high-flying promises of accuracy, speed, and automation in designing and delivering public services, from welfare to security (Veale and Brass 2019). The European Commission, for example, justifies its massive investments for AITs in health care and transportation as a matter of both green growth and the cost-effectiveness of public interventions (European Commission 2018).

- (2) At the same time, we observe severe disenchantment with AIT uses. This surfaces with highly publicized scandals around AITs in the public sector, such as for the Australian robodebt scheme⁴ or a Dutch welfare fraud detection system,⁵ but also successful grassroots pushbacks against ‘smart’ governance projects such as the *Toronto Sidewalk Lab* (Berger 2020). Critics of AIT techno solutionism emphasize not only severe risks of discrimination and other forms of rights violation of those exposed to data-driven decisions (Benjamin 2019; Eubanks 2018; Yeung 2019), but also point to the global injustices resulting from the vast energy and earthly mineral consumption and underpaid click work that go into AIT production (Crawford 2021). At the same time, it remains to be seen how seriously emergent initiatives for coercive regulation of public (and private) sector AIT uses – including in the European Union (European Commission 2021) – seek to and can constrain the business models and the sheer concentration of corporate power in the field of data-driven capitalism by companies, such as Amazon, Meta, or Google (Pasquale 2015; Zuboff 2019).

The insertion of AITs in the public realm – both as an alleged techno-solutionist tool in public decision-making *and* as a contested object for statutory regulation – should provoke a metaphorical call to scholarly arms for critical policy analysts. Indeed, the intellectual mission of the journal *Critical Policy Studies* (CPS) is to ‘challenge the limits of dominant positivist or techno-rationalist practices of policy analysis’, to bridge the gap between policy and practice, and ‘to bring politics back in’ with a focus on citizen participation, collaboration, and deliberative democracy (Griggs, Mathur, and Jas 2007, 1). It’s a wake-up call for CPS then when policymakers and some social scientists alike refer to AITs as a hyper-rational remedy for targeted, speedy, and accurate public decision-making (cf. for a wider debate: Coglianese and Lai 2021). It’s a wake-up call when regulators, including the EU, promote AIT as a green technology which will improve harvests or render public transport more energy-efficient (European Commission 2020) all while disavowing the *global* environmental and socio-economic implications of *producing* AITs from vast water and energy consumption, via rare mineral extraction, to underpaid click workers (Crawford 2021).

This forum seeks to stimulate the articulation of a critical research agenda on AI and public policy: What can critical policy studies contribute to the examination of AI and public policy? Bringing the initial CPS mission statement (Griggs, Mathur, and Jas 2007) to bear on the age of ‘smart’ governance, I identify three interconnected lines of critical inquiry: (1) interpretivist deconstruction of the norms and values that shape perceptions and uses of AI in public policy, including techno solutionism, (2) exploration of AITs in

public policy as a contingent practice of sociotechnical co-production, and (3) emancipatory critique of how existing smart governance projects and AIT regulation interact with (global) inequalities and power relations. The remainder of the article treats each of these in turn, synthesizing existing critical accounts, including from this journal, across disciplinary research niches.⁶ With this, I wish to contribute to the formation and consolidation of critically minded scholarship on public policy and artificial intelligence (we detail this research agenda in more depth in joint work with Emma Carmel and Jennifer Cobbe: Paul, Carmel, and Cobbe [forthcoming](#)).

Interpretivist deconstruction of techno-solutionist agendas

The first line of inquiry deconstructs actors' variable norms and judgment calls, and related struggles over meaning, around AI and public policy beyond any techno-solutionist claims. Since the 1970s, interpretivist methodology in policy studies has become an explicit antidote to a then dominant 'objectivist orientation to social reality [which had] failed to understand the subjective and intersubjective human dimensions' of policy problems and notions of social crisis (Fischer et al. 2015, 4). Techno-solutionism can be put in its place, from a CPS perspective, by exploring AI's interaction with policy as *semiotic artifacts* which result from selective representations of the social world, and the role of technology in it, and by investigating how such meaning-making conditions, and is conditioned by, specific political claims and power structures. I propose three specific analytical vantage points for this endeavor.

Firstly, we can explore policymakers' justifications for turning to AI and explore actors' ambiguous, contingent, and conflicting values, ideas, and motivations for supporting or resisting the insertion of AITs⁷ into the policy world, study how these meaning-making processes shape policies, and explore the structural conditions under which some interpretations of AIT become powerful in public policy contexts. In the techno-solutionist interpretation, AITs count as *objectifiers* of biased human decisions, as a technical fix for well agreed upon policy problems, and as cost-effectiveness booster in public administration. A wide-spread argument is that AI applications' alleged 'accuracy, consistency, speed, and capacity' would help public administrations overcome human bias (Coglianese and Lai 2021), and 'optimize' the public sector as a whole (Coglianese 2018). Leading regulators in the field argue that 'AI-enabled solutions can deliver shorter and richer feedback loops for all levels of governance, providing an opportunity to speed up, improve the efficiency and effectiveness of service delivery' (European Commission 2018, 19).

Arguably then, techno solutionism renders AITs as the high-tech version of Harold D. Lasswell's ideal policy analyst whose expertise informs the search for optimal solutions to politically pre-defined problems, thus assisting 'the realization of the value goals' with a scientifically neutral ethos (Lasswell 1970: 5). Such positivist accounts conflict with Science and Technology Studies' insight that technological choices are 'intrinsically political: they order society, distribute benefits and burdens, and channel power' (Jasanoff 2016, 243; on algorithmic ordering: Katzenbach and Ulbricht 2019). Techno-solutionist interpretations of AIT obviously also clash with CPS' interpretivist ontology of policymaking as a complex, highly selective, and contested process of vesting meanings – arguments, norms, knowledge claims, but also role ascriptions – in a 'policy' at the

discursive, epistemological, legal as well as practice level (cf., Carmel 2019; Fischer et al. 2015; Strassheim 2015). Consequently, to outsmart AI, CPS must reconstruct the role of techno solutionism *as constitutive part of* the politics of public policymaking.

Secondly, and relatedly, interpretivists should expose how and where value judgments come to bear in actual ‘smart’ governance projects, a task that does not lend itself to easy answers where boundaries between human and machine decisions are (presented as) nebulous. In her inaugural speech as White House Deputy Director for Science and Society, renowned sociologist Alondra Nelson stated on 19 May 2021: ‘when we provide inputs to the algorithm, when we program the device, when we design, test, and research, we are making *human* choices’. This concurs with constructivist understandings of algorithms, code, and data which insist that numerical representations of the world often involve neutralized value-based decisions about why and how we should count phenomena in the first place, what constitutes ‘data’ for what purposes, and how to use any analytical output (Coeckelbergh 2020: chapter 6; Gitelman 2013; Stone 2020).

Re-iterating the inevitability of human judgment as a CPS baseline for unpacking techno-solutionist promises remains urgent where (some) policymakers, policy analysts, and regulators attribute an ‘aura of accuracy and objectivity’ to algorithms and machines (Henman 2020, 213), when, meanwhile, evidence of ‘encoded’ racial, gender, and other forms of bias in AI-based decision-making accrues (Benjamin 2019; Noble 2018). Discriminatory effects may be obvious in highly scandalous *outputs* of AI-based systems, for example, when Google’s image recognition AI labels Black people as ‘gorillas’ (Hern 2018), or when the Dutch government wrongly accuses thousands of already marginalized immigrant families of welfare fraud based on AI recommendations (Henley and Booth 2020).

However, in most cases, interpretivist re-construction of judgment calls is more difficult. This is partly a technical problem as the encoded norms, coding rules, and selection mechanisms are ‘black-boxed’ and become ever less scrutable beyond the rare cases of highly obvious bias or error (Pasquale 2015; Yeung 2019). This is partly also because agency is ‘distributed’ in complex ways between (different) bits of code, hardware, and humans in algorithmic systems (Reddy, Cakici, and Ballesterero 2019), and because of the technical conflation of data gathering, analysis, action recommendation as well as the life updating of models for these calculative processes (more on these intricacies in Paul, Carmel, and Cobbe *forthcoming*). One related fear is that AITs intensify the technocratization and de-politicization of public decision-making. The CPS community must discuss the *methodical* implications of the obscurity of decision-making for exploring smart governance projects through interpretive methods.

Thirdly, CPS also helps treating key concepts for *regulating* AITs – and the related propositions for ‘ethical’ AIT design and usage – as politically inaugurated normative principles, which operate beyond neutral self-evidence and serve specific interests (cf. Paul, Carmel, and Cobbe *forthcoming*). STS scholars reject models of ‘evidence-based’ technology regulation, arguing that these misconceive technology as neutral, objectifiable, and exogenous to human value judgments about its uses, costs, and benefits (for nanotechnology, cf. Åm 2015). From an interpretivist perspective, concepts, such as ‘bias’, ‘accountability’, ‘transparency’, or ‘ethical AI’ cannot be taken for granted. Rather, they are contested semiotic constructs; activated or marginalized in *political* struggles over AIT regulation and smart governance projects.

For example, avoidance of ‘human bias’ counts as key motive in techno-solutionist explanations of the AI turn in public administration (Coglianese and Lai 2021), while, at the same time, the desire to mitigate ‘algorithmic bias’ drives positivist (and normative) calls for ‘risk-based’ AIT regulation (Krafft, Zweig, and König 2022). From a postpositivist perspective, attempts to ‘overcome’ biases by more and better data analyzed at more speed utterly miss the point. Some even suggest that the obsession with overcoming bias is a move of distraction by big tech industry to avoid meaningful regulation (Powles 2018).

Rather than obsessing about bias, we therefore require *critical debate* about (1) the framings of problems and solutions that get encoded in AIT regulation, and (2) the structures and power relations that render some of these interpretations policy-relevant while marginalizing others. Some research, including in this journal, duly politicizes technology regulation in that way. This shows, for example, how the EU’s General Data Protection Regulation’s framing of ‘risk’ prioritizes business interests over individual rights (Padden and Öjehag-Pettersson 2021) or how the EU’s emergent AI policy marginalizes questions of redistribution and economic inequality (Niklas and Dencik 2021).

Practice-oriented exploration of sociotechnical co-production ‘on the ground’

Of course, CPS scholarship does not stop at the reconstruction of discourses. As Frank Fischer suggested, policies on paper are mediated by ‘communicative practices’ in specific empirical settings on the ground and these mobilize or silence-specific sets of norms, arguments, and interpretations in contingent ways (Fischer et al. 2015, 5). This means that we need to explore specific use cases to reveal how explicit and implicit value statements on the use and regulation of AITs, and struggles over policy-relevant interpretations, actually bear on policy practice on the ground.

Firstly, the practice orientation for research comes with an imperative to study AIT-policy-interactions – and the related meaning-making processes – in its various articulations *below* the macro level of formal policy or regulatory programs (for the more general argument: cf., Carmel 2019; Dubois 2009). We are yet to consolidate critical policy-analytical insights on how the public sector *actually (and variably) incorporates* AITs and any abstract ideas about their functionality, promises, and risks in specific use cases. How are techno-solutionist promises for AIT in the policy world enacted through practices on the ground, by and for specific communities? How are they contested, disappointed, and potentially reworked, and by whom? How do different actors enact and contest abstract regulatory safeguards, such as ‘transparency’, or protection against ‘bias’ in real-life applications?

These questions remain rather unanswered in literatures which conceptualize AIT regulation either from an applied ethics perspective (e.g. the widely cited accounts on EU regulation: Goodman and Flaxman 2017; Wachter *et al.* 2017) or prescribe rational choice approaches to risk mitigation (e.g. Krafft, Zweig, and König 2022). Though coming from different methodological and disciplinary traditions, both proponents of ‘ethical’ AI and automation and those promoting ‘proportionate’ regulation thereof suggest that the highly variable contexts for using these technologies – from seemingly mundane chatbots all the

way to automated weapons – should shape the assessment of the risks and benefits of using the technology (Chatila 2021; Krafft, Zweig, and König 2022). From a CPS perspective, a shared limitation of these accounts – next to the idealist overstatement of the persuasive power of ethics in the former and the rational choice overstatement of the availability of ‘full’ knowledge about a technology and its risks in the latter – is their apolitical portraying of policies and regulation (I offer a systematic review of variable concepts of regulation in the AIT/automation literature in Paul [forthcoming](#)). There is little conceptual room for actors’ discretion over how to act upon the alleged promises and risks of AI in policy practice, and how to map them onto existing interests, institutions, and identities.

Perhaps the most comprehensive review of public sector AIT systems; a report on the 142 most significant Federal Agencies and Departments in the US (Engstrom et al. 2020); suggests that a change of perspectives is apt indeed. The study concludes that while 45% of these agencies have experimented with machine learning, there is a grave lack of technical capacity and knowledge within the administration that defies reflexive use and careful inspection of AI-based decision recommendations. Another study, on the appropriation of predictive technologies in the British police (Sandhu and Fussey 2021, 79) highlights ‘a range of fissures’ occurring as police officers consider the smart technologies they are expected to use flawed and resist AI’s intrusion of their discretion. Similarly, law and regulation scholars started grasping the varieties of conceptualizing and enacting ‘human control’ in man-machine systems across policy sectors and jurisdictions (Elish 2019; Firlej and Taeihagh 2021), highlighting how institutional norms, but also the features and limits of the technology itself, shape the regulatory meaning of ‘control’, ‘accountability’, or ‘liability’.

Secondly, with its focus on operationalizing human-machine interactions for regulation, the latter example raises an important focal point for researching practices of AI in public policy. It echoes a classic assumption in Science and Technology Studies that the development, use, and regulation of technology cannot be explored in the abstract because these processes are inexorably *situated* in specific sociotechnical contexts in which technology is *co-produced* with societal knowledge and institutions (Jasanoff 2004). Scholars in the CPS community translate the concept of sociotechnical co-production for policy contexts when stating that the ‘actionable’ or ‘performative’ features of policymaking materialize through complex, endogenic, and non-predictable interactions between human and non-human actors (including machines, but also animals, materials, infrastructure, or laboratory set-ups) (Wagenaar 2012). The methodological implication for critical analyses of AI-policy-interactions is that both technical and human affordances, restraints, and roles in policy practice – and dominant interpretations thereof – must enter the equation.

From this perspective, the wide-spread parlance of keeping ‘the human in the loop’ in any AIT application seems utterly counterproductive. It fathoms a neat distinction of human/machine decisions in smart governance settings, with humans ready to interfere when – and just when – the machine fails. Such dichotomy is not only *conceptually flawed*, as discussed above, it is also *democratically problematic* in the public sector, for at least two reasons:

- the focus on ‘the AIT’ as a machine actor releases publicly accountable actors from liability for choices on procuring and designing an AIT, selecting data, initial

coding, and programming, or how to incorporate the AI into hardware and how to use it in decision-making; choices which co-shape how the technology works in practice and how it affects citizens;

- the ‘human-in-the-loop’ model also turns policy experts and bureaucrats into ‘moral crumple zones’ who soak up liability for machine failure but have little discretion for professional, moral, or empathy-based judgment otherwise, and are being increasingly de-qualified in taking decisions without relying on ‘the machine’ (Elish 2019; cf. Yeung 2019). There is a risk that humans are degraded to mere correctors of alleged machine decisions but increasingly lack the knowledge and skills to actually do so.

Overall, the critical exploration of smart governance projects necessitates a move beyond dichotomous understandings of human-machine decisions (as well as related concepts like autonomy, cognition, or expertise). Practice-oriented policy research on AITs needs to tease out the politics of perceiving, regulating, and enacting interactions between humans and technology in concrete applications.

Emancipatory critique of the wider implications of the AIT-policy-complex

In the 2016 CPS forum *What is critical*, Frank Fischer urged the community to reach beyond ‘mere’ interpretivism and practice analysis. CPS must stay socially relevant, he argued, by situating our analyses of policy meanings and practice in a wider context of societal discourses and structures, and by working toward an emancipatory critique of power relations to induce social change (Fischer 2016). Critical inquiry involves attending to ‘the wider context and structural conditions in which governing is enacted, and with what implications for socio-political orders and inequalities in power’ (Carmel 2019, 41). Such critique can be *emancipatory*, Carmel argues, because ‘it can free us to see our world, ourselves and others, revealing privilege and oppression’ as a foundation for progressive change (ibid: 50).

To do so, firstly, CPS ought to situate the analysis of AI-related policy framings and practice against the abundant insights from ethical and socio-legal research on AITs’ effects on fundamental human rights, rule of law procedures in public decision contexts, and other normative baselines for democratic governance (Cobbe and Singh 2020; Coeckelbergh 2020; Dubber, Pasquale, and Das 2020; Mantelero 2018; Pasquale 2015; Veale and Brass 2019; Yeung 2019). Critical analysis of AITs and public policy must discuss such rights-related effects, especially where official policy and regulatory rhetoric frames the problem with AITs as matters of *individual* rights and marginalizes more structural forms of discrimination, domination, and inequality. This requires teasing out, for example, the overlooked ‘collective dimension’ of data protection concerns (Mantelero 2016), the lack of considering social rights in European AI policy (Niklas and Dencik 2021), but also the colonial repercussions of how the global North produces, regulates, and deploys AITs (Carmel and Paul 2022), including by extracting resources (e.g., labour, raw minerals) from, and externalising social and environmental costs to, the global South.

Several scholars have already situated their analyses of distinct AIT applications in the policy world against such wider patterns of exclusion and domination. For example,

analysts of ‘smart’ border policies highlight the risk of state authorities violating migrants’ fundamental rights, such as nondiscrimination or fair treatment without noticing these violations, being able to account for their occurrence, or correcting them in due course (Amoore and Hall 2009; Leese 2014). Scholars of predictive policing point to similar dynamics and discuss the self-fulfilling prophecies of algorithmic racial profiling, targeted interventions in ‘Black’ neighborhoods, ‘finding’ crime and feeding this ‘information’ back into an algorithm that has been trained with data that is racially biased (Kaufmann, Egbert, and Leese 2019; cf. wider debate in: Benjamin 2019). Social policy analysts equally worry about automated forms of discrimination which could deepen inequalities in access to social benefits and services (Eubanks 2018). Certainly, the specific forms of racialized discrimination in a residual welfare state might make the U.S. an extreme case. And yet: the Dutch welfare fraud detection scandal, which erroneously accused especially ethnic minority communities of fraud, but also uses of social credit systems (China) or mass biometric databases (India’s Aadhaar system) to control minority populations highlights the wider bearing of the AIT-driven acceleration of discrimination. These context-specific patterns of discrimination through AIT-policy-interactions require systematic unpacking.

Secondly, critical policy analysis of AITs also ought to explore whether and how policy meanings and policy practice reproduce more global power relations – explicitly or by omission. This requires our analyses to focus on the role of AITs and data-driven decision-making in socio-political ordering processes by reconstructing ‘the economic, cultural, and political contexts that both shape the design of algorithms as well as accommodate their operation’ (Katzenbach and Ulbricht 2019, 3; cf., Burrell and Fourcade 2021). At one level, this speaks to a tradition of poststructuralist *Ideologiekritik* of neoliberalism within CPS and other research communities. For example, the insertion of ‘learning to code’ in UK curricula can be discussed as a means of creating ‘governable citizens’ in the digital era (Williamson 2016), and the Norwegian digitalization agenda for schools as forging a neoliberal prototype of ‘future citizen-workers’ all while sidelining educational ideals of humanism and solidarity (Høydal and Haldar 2021). Foucauldians describe ‘disruptive’ adoptions of automation and AI technologies in the education sector, which serve to change the mind-sets of teachers and pupils alike and to condition their identities and activities alongside neo-liberal rationalities (Mertanen, Vainio, and Brunila 2021; Ramiel 2021).

At another level, to expose and critique more global forms of domination in a move of ‘*Herrschaftskritik*’ (Jessop and Sum 2016), critical analyses of AIT-policy-interactions can build on accounts of transformed state-corporate relations in the digital age. For example, social theorists describe a corporatist transformation of the state which increasingly focuses on ‘minting’ data and governing what is measurable, becomes an ‘open data portal’ also for private firms and a ‘provider of digital services’, and which thereby ‘opens itself up to competition from private alternatives’ in ways that undermine its democratic legitimacy (Fourcade and Gordon 2020, 95). This transformation might accelerate through decision-makers’ and regulators’ strategic prioritization of competitiveness goals in the global ‘race to AI’ (cf., Paul forthcoming). Even gloomier accounts suggest the digital age will eventually obliterate territorial forms of state organization altogether, while enabling (the often corporate) ‘data controllers’ both to monetize data/patterns and to determine future access to private and public services (Pistor 2020).

Systematic explorations of the interactions of AIT uses in public policy with the shifting power relations between big tech, states, and citizens are still rare. Research in the migration domain suggests that public uses of AITs create profitable business cases for technology development (Carmel 2017; Leese 2018) and deliberately provide tech firms with existential ‘testing grounds’ for their innovations (Molnar 2021). A case study on Swedish social services provision identifies an equally constitutive amalgamation of business interests and public sector control ambitions through automation projects (Germundsson 2022). More research along these lines will be crucial for pinning down critique of corporatist state transformation.

Such debates also throw an anchor for the *emancipatory agenda* that CPS holds dear since its inception: the formulation of democratic and progressive alternatives. We observe, among others, advances in democratic theory to reflect upon the use of AITs in community-based forms of political resistance (Crawford et al. 2013), interventionist conceptualizations of ‘data justice’ to counter the local and global inequalities coming with ‘datafied’ societies and welfare states (Dencik et al. 2019), or calls for the creation of new ‘abolitionist tools’ to overcome encoded forms of bias of our age (Benjamin 2019). Some even call for ‘digital socialism’ that involves citizens and consumers in data-driven decision-making – public and private – in the design, data selection, modeling of use cases, application and control of AITs and other data-driven technologies (Fourcade and Gordon 2020).

At the very least, the research agenda promoted here – with its triple focus on (1) the interpretivist reconstruction of values and meanings in AIT-policy-interactions, (2) the exploration of concrete practices of sociotechnical co-production, and (3) critical discussion of how AIT uses in policy interact with (global) inequality and wider power relations – seeks to set a scholarly counterpoint to techno solutionism. Beyond ‘out-smarting’ AI in academic debate, however, critically minded scholars of AI and public policy face a deeply political question: ‘how might these technologies be governed in such a way that they better and more equitably serve the public?’ (Fourcade and Gordon 2020, 96). Let critical inquiry on AI and public policy bring on this crucial debate!

Notes

1. As smart policy expert Paul Henman writes ‘AI is typically used to refer to (technological systems developed with) machine learning algorithms that self-organise their internal variables and values to achieve desired outcomes’; in a nutshell: to ‘computer systems built (at least partially) using machine learning’ (Henman 2020, 210). In our continuous discussions, Emma Carmel has rightly pointed out the importance of the “T” in AITs: the complex contingency of how AI is inserted into hardware and software requires treating them as a set of highly variable AI *technologies*.
2. Henman discusses these three functions in more depth (2020: 210–213).
3. i.e. the analysis of large amounts and formerly unrelated points of data at high speed, based on enormous calculative capacities of (ever growing) IT infrastructures, with the aim of identifying correlations and patterns as well as predicting future behavior.
4. Using an algorithm-supported method of automated debt assessment and recovery, the Australian authorities unlawfully raised \$ 1.76 billion in debts against more than 440.000 people between 2015 and 2019 (Henriques-Gomes 2021).
5. The Dutch scheme received criticism not only because of high error rates but also because the fraud detection algorithm disproportionately ‘identified’ welfare recipients of ethnic

minority background as fraudulent. A court ruled the system unlawful and discriminatory in 2021 (Henley and Booth 2020).

6. Many of the critical voices in the domain appear in technology-focused journals such as *Big Data & Society*; *Computer Law & Security Review*; *Ethics and Information Technology*; *Information, Communication & Society*; *Minds and Machines*; *Science and Engineering Ethics*; or *Science, Technology & Human Values*. Their lack of foundation in critical policy studies mean, however, that implications of ‘AI turns’ for how we understand and study public policies and are only partially addressed.
7. This should also include analysis of merely rhetorical turns to AI as a ‘moniker’ that could be ‘more marketing lingo than technical description’ (Henman 2020: 217).

Acknowledgements

I acknowledge highly stimulating debates with Emma Carmel and Jennifer Cobbe around our forthcoming co-edited Handbook on Public Policy and Artificial Intelligence. Whilst this forum carves out my own (limited) critical policy studies view on AIT, the Handbook’s introduction develops a critical approach on a much larger canvas thanks to Emma and Jennifer’s expertise. I thank Heidrun Åm, two anonymous reviewers, as well as forum editors Kathrin Braun and Dieter Plehwe for helpful feedback on draft versions.

Disclosure statement

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

Notes on contributor

Regine Paul is Professor at the Department of Government, University of Bergen, Norway. Her research explores the critical political economy of migration governance, risk regulation, and artificial intelligence regulation in Europe, but also the political uses of risk analysis heuristics in public decision-making. She co-edits the forthcoming *Handbook on Public Policy and Artificial Intelligence* (Elgar) and *Critical Policy Studies*.

References

- Åm, H. 2015. “Co-production and Public Policy: Evidence, Uncertainty and socio-materiality.” In *Handbook of Critical Policy Studies*, edited by F. Fischer, et al., 297–316. Cheltenham Spa: Edward Elgar.
- Amoore, L., and A. Hall. 2009. “Taking People Apart: Digitised Dissection and the Body at the Border.” *Environment and Planning D: Society and Space* 27 (3): 444–464. doi:10.1068/d1208.
- Benjamin, R. 2019. *Race after Technology: Abolitionist Tools for the New Jim Code*. Medford, MA: Polity.
- Berger, B. 2020. “Sidewalk Labs’ Failure and the Future of Smart Cities.” *Triple Pundit*. www.triplepundit.com.
- Burrell, J., and M. Fourcade. 2021. “The Society of Algorithms.” *Annual Review of Sociology* 47: 213–237. doi:10.1146/annurev-soc-090820-020800.
- Carmel, E. 2017. “Re-interpreting Knowledge, Expertise and EU Governance: The Cases of Social Policy and Security Research Policy.” *Comparative European Politics* 15 (5): 771–793. doi:10.1057/s41295-016-0079-1.
- Carmel, E. 2019. “Regimes of Governing Practices, socio-political Order and Contestation.” In *Governance Analysis: Critical Enquiry at the Intersection of Politics, Policy and Society*, edited by E. Carmel, 24–46. Cheltenham: Edward Elgar.

- Carmel, E., and R. Paul. 2022. "Peace and Prosperity for the Digital Age? The Colonial Political Economy of European AI Governance." *IEEE Technology and Society Magazine* 41 (2). doi:10.1109/MTS.2022.3173340.
- Cath, C., S. Wachter, B. Mittelstadt, M. Taddeo, and L. Floridi. 2018. "Artificial Intelligence and the "Good Society": The US, EU, and UK Approach." *Science and Engineering Ethics* 24 (2): 505–528. doi:10.1007/s11948-017-9901-7.
- Chatila, R. 2021. "Trustworthy AI." In *Reflections on Artificial Intelligence for Humanity*, edited by B. Braunschweig, M. Ghallab, et al., 13–39. Cham: Springer International Publishing.
- Cobbe, J., and J. Singh. 2020. "Reviewable Automated Decision-Making." *Computer Law & Security Review* 39: 105475. doi:10.1016/j.clsr.2020.105475.
- Coeckelbergh, M. 2020. *AI Ethics*. Cambridge, MA: MIT Press.
- Coglianese, C. 2018. "Optimizing Regulation for an Optimizing Economy." Rochester, NY: Social Science Research Network, (accessed September 2021). available at: <https://papers.ssrn.com/abstract=3300304>
- Coglianese, C., and A. Lai 2021. "Algorithm V. Algorithm: A Framework for Governmental Use of Machine Learning." In *ECPR General Conference*, digital conference. 2021.
- Crawford, K., et al. 2013. "Big Data, Communities and Ethical Resilience: A Framework for Action." <https://www.rockefellerfoundation.org/wp-content/uploads/71b4c457-cdb7-47ec-81a9-a617c956e6af.pdf>
- Crawford, K. 2021. *Atlas of AI: The Real Worlds of Artificial Intelligence*. New Haven: Yale University Press.
- Dencik, L., A. Hintz, J. Redden, and E. Treré. 2019. "Exploring Data Justice: Conceptions, Applications and Directions." *Information, Communication & Society* 22 (7): 873–881. doi:10.1080/1369118X.2019.1606268.
- Dubber, M. D., F. Pasquale, and S. Das, (edited by). 2020. *The Oxford Handbook of Ethics of AI*. Oxford: Oxford University Press.
- Dubois, V. 2009. "Towards a Critical Policy Ethnography: Lessons from Fieldwork on Welfare Control in France." *Critical Policy Studies* 3 (2): 221–239. doi:10.1080/19460170903385684.
- Elish, M. C. 2019. "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction (pre-print)". *SSRN Research Papers*, doi:10.2139/ssrn.2757236.
- Engstrom, D. F., D. E. Ho, C. M. Sharkey, and M.-F. Cuéllar 2020. "Government by Algorithm." *Artificial Intelligence in Federal Administrative Agencies*, Stanford, Calif.: Administrative Conference of the United States.
- Eubanks, V. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.
- European Commission. 2018. *Coordinated Plan on Artificial Intelligence*. Brussels: EC.
- European Commission. 2020. "White Paper on Artificial Intelligence - A European Approach to Excellence and Trust." Brussels.
- European Commission. 2021. *Proposal for a Regulation Laying down Harmonised Rules on Artificial Intelligence and Amending Certain Union Legislative Acts*. Brussels.
- Firlej, M., and A. Taeihagh. 2021. "Regulating Human Control over Autonomous Systems". *Regulation & Governance* first online. 15 1071–1091. 10.1111/rego.12344.
- Fischer, F. 2016. "What Is Critical? Connecting the Policy Analysis to Political Critique." *Critical Policy Studies* 10 (1): 95–98. doi:10.1080/19460171.2015.1129350.
- Fischer, F., D. Torgerson, A. Durnová, and M. Orsini. 2015. "Introduction to Critical Policy Studies." In *Handbook of Critical Policy Studies*, edited by F. Fischer, et al., 1–24. Cheltenham Spa: Edward Elgar.
- Fourcade, M., and J. Gordon. 2020. "Learning like a State: Statecraft in the Digital Age." *Journal of Law and Political Economy* 1 (1): 78–108. doi:10.5070/LP61150258.
- Germundsson, N. 2022. "Promoting the Digital Future: The Construction of Digital Automation in Swedish Policy Discourse on Social Assistance." *Critical Policy Studies* first online) 1–19. 10.1080/19460171.2021.2022507.
- Gitelman, L., ed. 2013. *'Raw Data' Is an Oxymoron*. Cambridge, MA: MIT Press.

- Goodman, B., and S. Flaxman. (2017). European Union Regulations on Algorithmic Decision-Making and a “Right to Explanation”. *AIMag*, 38(3), 50–57. [10.1609/aimag.v38i3.2741](https://doi.org/10.1609/aimag.v38i3.2741)
- Griggs, S., N. Mathur, and P. Jas. 2007. “Editorial.” *Critical Policy Studies* 1 (1): 1–1. doi:[10.1080/19460171.2007.9518505](https://doi.org/10.1080/19460171.2007.9518505).
- Henley, J., and R. Booth (2020). “Welfare Surveillance System Violates Human Rights, Dutch Court Rules.” *The Guardian* 5 February (accessed December 2021).
- Henman, P. 2020. “Improving Public Services Using Artificial Intelligence: Possibilities, Pitfalls, Governance.” *Asia Pacific Journal of Public Administration* 42 (4): 209–221. doi:[10.1080/23276665.2020.1816188](https://doi.org/10.1080/23276665.2020.1816188).
- Henriques-Gomes, L. (2021) “Robodebt: Court Approves \$1.8bn Settlement for Victims of Government’s ‘Shameful’ Failure’.” *The Guardian* 11 June (accessed March 2022).
- Hern, A. 2018. “Google’s Solution to Accidental Algorithmic Racism: Ban Gorillas.” *The Guardian* 12 January (accessed December 2021).
- Høydal, Ø. S., and M. Haldar. 2021. “A Tale of the Digital Future: Analyzing the Digitalization of the Norwegian Education System’.” *Critical Policy Studies* first online 1–18. [10.1080/19460171.2021.1982397](https://doi.org/10.1080/19460171.2021.1982397).
- Jasanoff, S., ed. 2004. *States of Knowledge: The Co-production of Science and the Social Order*. 1 ed. London: Routledge.
- Jasanoff, S. 2016. *The Ethics of Invention: Technology and the Human Future*. New York: W.W. Norton & .
- Jeandesboz, J. 2021. “Technology, Knowledge and the Governing of Migration.” In *Handbook on the Governance and Politics of Migration*, edited by E. Carmel, K. Lenner, and R. Paul, 329–340. Cheltenham: Edward Elgar .
- Jessop, B., and N.-L. Sum. 2016. “What Is Critical?” *Critical Policy Studies* 10 (1): 105–109. doi:[10.1080/19460171.2015.1129352](https://doi.org/10.1080/19460171.2015.1129352).
- Katzenbach, C., and L. Ulbricht. 2019. “Algorithmic Governance.” *Internet Policy Review* 8 (4). doi:[10.14763/2019.4.1424](https://doi.org/10.14763/2019.4.1424).
- Kaufmann, M., S. Egbert, and M. Leese. 2019. “Predictive Policing and the Politics of Patterns.” *The British Journal of Criminology* 59 (3): 674–692. doi:[10.1093/bjc/azy060](https://doi.org/10.1093/bjc/azy060).
- Krafft, T. D., K. A. Zweig, and P. D. König. 2022. “How to Regulate Algorithmic decision-making: A Framework of Regulatory Requirements for Different Applications.” *Regulation & Governance* 16 (1): 119–136. doi:[10.1111/rego.12369](https://doi.org/10.1111/rego.12369).
- Lasswell, H. D. 1970. “The Emerging Conception of the Policy Sciences.” *Policy sciences* 1: 3–14.
- Leese, M. 2014. “The New Profiling: Algorithms, Black Boxes, and the Failure of anti-discriminatory Safeguards in the European Union.” *Security Dialogue* 45 (5): 494–511. doi:[10.1177/0967010614544204](https://doi.org/10.1177/0967010614544204).
- Leese, M. 2018. “Standardizing Security: The Business Case Politics of Borders.” *Mobilities* 13 (2): 261–275. doi:[10.1080/17450101.2017.1403777](https://doi.org/10.1080/17450101.2017.1403777).
- Mantelero, A. 2016. “Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection.” *Computer Law & Security Review* 32 (2): 238–255. doi:[10.1016/j.clsr.2016.01.014](https://doi.org/10.1016/j.clsr.2016.01.014).
- Mantelero, A. 2018. “AI and Big Data: A Blueprint for A Human Rights, Social and Ethical Impact Assessment.” *Computer Law & Security Review* 34 (4): 754–772. doi:[10.1016/j.clsr.2018.05.017](https://doi.org/10.1016/j.clsr.2018.05.017).
- Mertanen, K., S. Vainio, and K. Brunila. 2021. “Educating for the Future? Mapping the Emerging Lines of Precision Education Governance.” *Policy Futures in Education*.
- Molnar, P. . 2021. “Technological Testing Grounds. Migration Management Experiments and Reflections from the Ground Up.” EDRi and Refugee Law Lab
- Morozov, E. 2014. *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: PublicAffairs.
- Niklas, J., and L. Dencik. 2021. “What Rights Matter? Examining the Place of Social Rights in the EU’s Artificial Intelligence Policy Debate.” *Internet Policy Review* 10 (3). doi:[10.14763/2021.3.1579](https://doi.org/10.14763/2021.3.1579).
- Noble, S. U. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. Illustrated ed. New York: New York University Press.

- Padden, M., and A. Öjehag-Pettersson. 2021. “Protected How? Problem Representations of Risk in the General Data Protection Regulation (GDPR).” *Critical Policy Studies* 15 (4) : 486–503.
- Pasquale, F. 2015. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge: Harvard University Press.
- Paul, R. *forthcoming*. “The Politics of Regulating Artificial Intelligence Technology: A Competition State Perspective.” In *Handbook on Public Policy and Artificial Intelligence*, edited by R. Paul, E. Carmel, and J. Cobbe, 1-33. Cheltenham: Edward Elgar.
- Paul, R., E. Carmel, and J. Cobbe. *forthcoming*. “Public Policy and Artificial Intelligence: Vantage Points for Critical Inquiry.” In *Handbook on Public Policy and Artificial Intelligence*, edited by R. Paul, E. Carmel, and J. Cobbe, 1-24. Cheltenham: Edward Elgar.
- Pistor, K. 2020. “Statehood in the Digital age1.” *Constellations* 27 (1): 3–18. doi:10.1111/1467-8675.12475.
- Powles, J. 2018. “The Seductive Diversion of ‘Solving’ Bias in Artificial Intelligence.” *OneZero (Blog)*, posted 7 Dec.
- Prainsack, B. 2020. “The Value of Healthcare Data: To Nudge, or Not?” *Policy Studies* 41 (5): 547–562. doi:10.1080/01442872.2020.1723517.
- Ramiel, H. 2021. “Edtech Disruption Logic and Policy Work: The Case of an Israeli Edtech Unit.” *Learning, Media and Technology* 46 (1): 20–32. doi:10.1080/17439884.2020.1737110.
- Reddy, E., B. Cakici, and A. Ballestero. 2019. “Beyond Mystery: Putting Algorithmic Accountability in Context.” *Big Data & Society* 6 (1): 205395171982685. doi:10.1177/2053951719826856.
- Sandhu, A., and P. Fussey. 2021. “The “Uberization of Policing”? How Police Negotiate and Operationalise Predictive Policing Technology.” *Policing and Society* 31 (1): 66–81. doi:10.1080/10439463.2020.1803315.
- Stone, D. 2020. *Counting: How We Use Numbers to Decide What Matters*. New York: Norton.
- Strassheim, H. 2015. “Politics and Policy Expertise: Towards a Political Epistemology.” In *Handbook of Critical Policy Studies*, edited by F. Fischer, et al., 319–340. Cheltenham Spa: Edward Elgar.
- Ulnicane, I., W. Knight, T. Leach, B. C. Stahl, and W.-G. Wanjiku. 2020. “Framing Governance for a Contested Emerging Technology: Insights from AI Policy.” *Policy and Society* first online): 1–20.
- Veale, M., and I. Brass. 2019. “Administration by Algorithm? Public Management Meets Public Sector Machine Learning.” In *Algorithmic Regulation*, edited by K. Yeung and M. Lodge, 1-30. Oxford: Oxford University Press.
- Wachter, S., B. Mittelstadt, and L. Floridi. 2017. “Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation.” *International Data Privacy Law* 7 (2): 76–99. doi:10.1093/idpl/ix005.
- Wagenaar, H. 2012. “Dwellers on the Threshold of Practice: The Interpretivism of Bevir and Rhodes.” *Critical Policy Studies* 6 (1): 85–99. doi:10.1080/19460171.2012.659890.
- Williamson, B. 2016. “Political Computational Thinking: Policy Networks, Digital Governance and “Learning to Code”.” *Critical Policy Studies* 10 (1): 39–58. doi:10.1080/19460171.2015.1052003.
- Yeung, K. 2019. “Why Worry about decision-making by Machine?” In *Algorithmic Regulation*, edited by K. Yeung and M. Lodge, 21–48. Oxford: Oxford University Press.
- Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.