

How empty is Trustworthy AI? : A discourse analysis of the Ethics Guidelines of Trustworthy AI

Authors	Stamboliev,Eugenia; Christiaens,Tim
Published in	Critical Policy Studies
DOI	10.1080/19460171.2024.2315431
Publication Date	2024-02-11
Document Version	publishersversion
Link	https://research.tilburguniversity.edu/en/publications/4a2bc71a-4e58-4d65-a063-6865049dd7d8
Citation	Stamboliev, E & Christiaens, T 2024, 'How empty is Trustworthy AI? : A discourse analysis of the Ethics Guidelines of Trustworthy AI', Critical Policy Studies, vol. 19, no. 1, pp. 39-556. https://doi.org/10.1080/19460171.2024.2315431
Download Date	2026-05-17 12:42:48
Rights	<p>General rights</p> <p>Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.</p> <ul style="list-style-type: none"> - Users may download and print one copy of any publication from the public portal for the purpose of private study or research. - You may not further distribute the material or use it for any profit-making activity or commercial gain - You may freely distribute the URL identifying the publication in the public portal" <p>Take down policy</p> <p>If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.</p>

How *empty* is Trustworthy AI? A discourse analysis of the Ethics Guidelines of Trustworthy AI

Eugenia Stamboliev & Tim Christiaens

To cite this article: Eugenia Stamboliev & Tim Christiaens (11 Feb 2024): How *empty* is Trustworthy AI? A discourse analysis of the Ethics Guidelines of Trustworthy AI, Critical Policy Studies, DOI: [10.1080/19460171.2024.2315431](https://doi.org/10.1080/19460171.2024.2315431)

To link to this article: <https://doi.org/10.1080/19460171.2024.2315431>



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



Published online: 11 Feb 2024.



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



View related articles [↗](#)



View Crossmark data [↗](#)

How *empty* is Trustworthy AI? A discourse analysis of the Ethics Guidelines of Trustworthy AI

Eugenia Stamboliev^a and Tim Christiaens^b

^aDepartment of Philosophy, University of Vienna, Vienna, Austria; ^bDepartment of Philosophy, Tilburg University, Tilburg, Netherlands

ABSTRACT

'Trustworthy artificial intelligence' (TAI) is contested. Considering the growing power of Big Tech and the fear that AI ethics lacks sufficient institutional backing to enforce its norms on AI industry, we struggle to reconcile ethical and economic demands in AI development. To establish such a convergence in the European context, the European Commission published the *Ethics Guidelines for Trustworthy AI* (EGTAI), aiming to strengthen the ethical authority and find common ground among AI industry, ethicists, and legal regulators. At first glance, this attempt allows to unify different camps around AI development, but we question this unity as one that subordinates the ethical perspective to industry interests. By employing Laclau's work on empty signifiers and critical discourse analysis, we argue that the EU's efforts are not pointless but establish a chain of equivalences among different stakeholders by promoting 'TAI' as a unifying signifier, left open so that diverse stakeholders unite their aspirations in a common regulatory framework. However, through a close reading of the EGTAI, we identify a hegemony of AI industry demands over ethics. This leaves AI ethics for the uncomfortable choice of affirming industry's hegemonic position, undermining the purpose of *ethics* guidelines, or contesting industry hegemony.

KEYWORDS

Trustworthy AI; policy analysis; Laclau; AI ethics; empty signifier; hegemony

1. Introduction

With new developments in Artificial Intelligence (AI)¹ have come ethical controversies surrounding its regulation. AI-systems have been accused of, among others, manipulating political deliberation on social media (Davies 2018; Sunstein 2017), reproducing human biases in algorithmic decision-making (Benjamin 2019; Edelman, Luca, and Svirsky 2017; Eubanks 2019), increasing humankind's ecological footprint (Coeckelbergh 2021; Crawford 2021). Cathy O'Neil (2016) even identifies AI tools as 'weapons of math destruction'. Considering the growing power of Big Tech (Bartoletti 2020), the design and regulation of AI have become precarious balancing acts between ethical and industry demands. The institutions wishing to promote ethical uses of AI

CONTACT Eugenia Stamboliev  eugenia.stamboliev@unvie.ac.at  Department of Philosophy, University of Vienna, NIG, Universitätsstraße 7, Vienna 1010, Austria

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

This is an Open Access article distributed under the terms of the Creative Commons Attribution-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. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

often lose the battle against industry interests, which induces Luke Munn (2022) to call AI ethics ultimately useless. The critics highlight that ethical principles in AI regulation often formulate non-binding utterances involved in ethics washing: AI industry simply buys ethical certification when suitable, rather than earning ethical approval with responsible business practices. Reinhardt (2022) points to conceptual issues with ethics labels like ‘trust’ or ‘trustworthy AI’ (TAI). For instance, ‘guidelines differ in the envisioned addressee of trust building’ (Reinhardt 2022, 2). Sometimes the general public is considered the main addressee of trust-building initiatives, but often AI regulations address themselves almost exclusively to corporate agents. Furthermore, ‘trust in AI’ and ‘TAI’ are often used synonymously to confusing effect (Ryan 2020). The discourse of trustworthiness becomes ‘a land of plenty’ (Reinhardt 2022, 4) with too many meanings lacking agreement on what counts for whom, which ultimately erodes ethics’ efficacy as a regulatory agent. Because of this conceptual vagueness and the lack of powerful enforcement, ethics fails to assert itself vis-à-vis industry interests, despite Mittelstadt (2019) counting more than 80 frameworks on AI ethics.

In this paper, we examine the authority of AI ethics (or the lack thereof) in the ‘TAI’-framework promoted by the European Commission and the High-Level Expert Group (HLEG) in its ‘Ethics Guidelines for Trustworthy AI’ (EGTAI). While the EU has in recent years developed multiple legislative initiatives for AI regulation (AI Act, Digital Services Act (DSA), Platform Work Directive), the EGTAI formulated an original inspiration for these projects. By studying the EU’s regulatory intentions at their inceptions, we hope to uncover the original tensions of AI regulation in their clearest state. Through a critical discourse analysis of these guidelines and the accompanying policy recommendations (HLEG 2019a, 2019b), we argue that AI ethics is not toothless or useless, but engaged in a struggle for hegemony with AI industry. Consistent with the program of critical policy analysis, we argue to ‘bring politics back in’ policy studies (Paul 2022, 498). As Paul highlights in her research agenda for critical policy studies in AI development, key concepts for regulating AI sometimes play a covert political role in technocratic and seemingly neutral policy discourses (Paul 2022, 500). More specifically, we employ Ernesto Laclau’s and Chantal Mouffe’s political philosophy of signification to show that ‘TAI’ operates as an empty signifier in the EGTAI. ‘TAI’ is not meant to have a fixed or stable signification, but to align multiple stakeholders into a political coalition. In the case of EGTAI (HLEG 2019a, 2), the document focuses on three pillars, each represented by a particular stakeholder. Establishing ‘TAI’ requires that AI is (a) lawful, represented by legal experts, (b) ethical, represented by individuals who professionally evaluate AI from the perspective of moral and social values, like academic ethicists independent from industry, and (c) robust, represented by AI industry and its technical experts. Promoting ‘TAI’ as an overarching principle allows the European Commission to bring these different groups and their divergent demands together into a single regulatory project. ‘TAI’ needs to be a ‘land of plenty’ to accommodate for these divergent and potentially conflicting viewpoints.

To avoid confusion, identifying ‘TAI’ as an empty signifier, firstly, does not imply that ‘TAI’ is nonsensical. It rather serves the political purpose of aligning different stakeholders in a common regulatory framework. Nonetheless, the critics are right in highlighting how the EGTAI surreptitiously favors AI industry interests over ethical concerns. Within the ‘TAI’-framework, the EGTAI, indeed, grants a hegemonic position

to AI industry vis-à-vis AI ethics. This is especially ironic given that the EGTAI explicitly promotes *ethics* guidelines for AI industry. This tension leads to a dilemma, which we will discuss in the final section of the paper: either AI ethics is reduced a mere symbolic seal of approval for industry initiatives, or we affirm a fundamental antagonism between ethical and industry demands, which can only be resolved through political conflict. Second, we treat ‘ethics’ and ‘industry’ as two distinct and identifiable camps, despite real-life potential overlaps and fuzzy boundaries. This generalization follows the EGTAI’s own assumption that both camps are clearly distinct ‘pillars’. While ‘AI industry’ generally refers to Big Tech companies, there are distinct nonprofit and research organizations that primarily focus on advancing ethical norms to constrain business interests. The added advantage of maintaining the EGTAI’s vocabulary is that it clearly emphasizes the political stakes of AI regulation (Coeckelbergh 2020; Munn 2022). We argue that the necessity for the EGTAI arises from the sensed urgency of tensions between these opposing forces and from the need for establishing harmonizing principles to pacify the use of AI in business applications. The clash between ethics and industry is of paramount importance to the EGTAI, even if it may not always be possible in practice to clearly separate the two or identify specific companies or ethical organizations aligned with either of both camps.

2. Empty signifiers and hegemony in critical discourse analysis

A helpful starting point for critical discourse analysis is the distinction between politics and the political that Laclau and Mouffe regularly invoke (Laclau 1996, 84; Mouffe 2005, 8–9; 2013, xii). Political science focuses on politics, i.e. the official institutions responsible for policymaking, like parliaments or executives. When one, for instance, studies policy-documents like the EGTAI, political science investigates the main institutional actors, their mutual relations, and official decision-making procedures.² The political, however, denotes fundamental conflicts of value that permeate an entire society. The political divides populations into opposing sides with contradictory perspectives struggling for dominance. Constitutive of the social are, in other words, insurmountable conflicts that take place not just in established institutions but also on the streets, the internet, or in high-level expert groups. Animating the policy-making procedures of, for instance, EU institutions are fundamentally different and opposed viewpoints on the European community’s identity. As Laclau (2005, 80) aptly summarizes, ‘the social is but the locus of irreducible tension’. Making political antagonism fundamental to social life, however, has implications for social analysis. If conflict over collective identity is inevitable, then no society can ever be said to have a single, positively defined collective identity. There is no objective reality to bind all members of a community together in a single political unit. Different groups or ‘subjects’ (Laclau and Mouffe 2014, 101) have conflicting views on collective identity and divergent policy demands.

The political unity lacking in objective reality subsequently must be constructed on the level of discourse (Marchart 2007, 136; Colpani 2022, 228). Societies might not be able to *naturally* constitute a homogenous whole, but they can still *represent* themselves as a political unity (Laclau 1996, 42). According to Laclau and Mouffe (2014, 98–99), ‘if the social does not manage to fix itself in the intelligible and instituted forms of a society, the social only exists, however, as an effort to construct that

impossible object. Any discourse is constituted as an attempt to dominate the field of discursivity, to arrest the flow of differences, to construct a centre'. Social unity is constructed through discursive means. Different subjects articulate discourses that represent *their* views as being those of society. They use rhetorical artifice to make one part of society stand in for the whole (Laclau 2005, 72). The universal consensus over collective identity, which does not objectively exist, is staged by making one particular representation of this identity stand in for the universal (Laclau 1996, 15). However, if multiple yet particular subjects simultaneously claim to represent the whole, social conflict returns on the discursive level (Colpani 2021, 229; Marchart 2007, 140), struggling to construct 'society' as a homogeneous whole according to their own particular demands. This is not a mere conflict of interest but a conflict over the tenets of collective identity, what 'the people' ultimately stands for. In debates over climate change, for example, ecomodernists stress a shared faith in technological progress, while green growth enthusiasts support sustainable economic growth, and degrowth ecologists want to scale back economic activity entirely. This disagreement cannot be settled through neutral scientific debate. It requires a more fundamental, democratic struggle over green politics. Laclau and Mouffe call this 'struggles for hegemony' (HLEG 2019a, 39). Diverse subjects with heterogeneous demands conflict over the meaning and purpose of the collective, which cannot be resolved through political bargaining or neutral scientific assessment. These distinct hegemonic projects are mutually incompatible. Both parties must struggle to determine the overall course of society's self-understanding and policy-making and any resulting hegemonic formation is irretrievably unstable and fragile as contestation can always reemerge (Laclau and Mouffe 2014, 135).

According to Laclau (2005, 78), subjects attempt to solidify their hegemonic projects by constructing 'chains of equivalences' with other political subjects. They articulate discourses that represent multiple different demands as equivalent to justify political coalitions between different subjects (Howarth 2010, 318). In climate policy for example, ecomodernists can establish a chain of equivalences between techno-utopian entrepreneurs and green growth supporters by articulating a discourse that links financial investments in tech innovation to overcoming fossil-fuel-based growth. Combatting climate change is thereby framed as not just an opportunity for ecomodernist entrepreneurs but also as a vehicle for green growth. Discursively established chains of equivalences thus mobilize different subjects for the same hegemonic project by representing one demand as unifying the demands of other subjects as well. Among the different constituent subjects of a chain of equivalences, one is thus hegemonic. It dictates the overall agenda, but needs allies to render the hegemonic project politically viable (Laclau and Mouffe 2014, 106). This leading subject represents the *universal* position within the hegemonic project – it purportedly speaks for all constituent subjects – whereas other groups only represent *particularist* perspectives within the larger whole. So long as the constituent subjects of a hegemonic project agree on this chain of equivalences, a politically stable ensemble of subjects is secured. Different hegemonic projects – with each their own internal coalitions and divisions, universalities and particularities – subsequently struggle for dominance over the social and political institutions responsible for policy-making. This entails participation in elections and formal democratic procedures in the realm of politics, but also informal practices like organizing bottom-up social

movements or infiltrating the bureaucracies of governmental institutions (Mouffe 2018, 19–20). These projects succeed insofar as they convince the population to regard *their* discursive representation of collective identity as ‘common sense’, i.e. the presumably self-evident and unquestioned framework with which everyday reality is interpreted (Brevini 2021, 145; Laclau and Mouffe 2014, 147).

Laclau and Mouffe’s critical discourse analysis hence focuses on how discourses establish chains of equivalences between different subjects to build hegemonic projects. This will prove helpful when studying the EGTAI insofar as the latter explicitly tries to harmonize the demands of three constituent subjects, namely legal experts, AI ethicists, and AI industry. The most important tactic in the hegemonic toolbox, for Laclau (2005, 71), is the articulation of ‘empty signifiers’ (Rear and Jones 2013, 376). These are not meaningless or nonsensical words, but signifiers not strictly tied to a particular signified. These are words or symbols of which the meaning is emptied out in order to accommodate for divergent viewpoints and interpretations. Their meaning is open-ended, covering up deep disagreements behind a discursively constructed consensus. Laclau specifically argues that the supposed ‘vagueness’ of empty signifiers ‘is not the result of any ideological or political underdevelopment; it simply expresses the fact that any populist unification takes place on a radically heterogeneous social terrain’ (Laclau 2005, 98). Empty signifiers are politically useful to affectively unify divergent subjects into a political coalition. Ecologists might, for example, struggle for ‘sustainability’ without specifying what ‘sustainability’ means or entails (Brown 2016; Swyngedouw 2010). For green growth supporters, this could mean investing in energy-saving technology, whereas degrowth activists want to reduce energy consumption entirely. As empty signifier ‘sustainability’ binds both viewpoints together in a single political coalition. Ultimately, there is always a hegemonic subject within the chain of equivalence able to put forward its own particular interpretation of ‘sustainability’ as the universal position, but it has to disavow this particularist origin in order to incorporate the other constituent subjects. A green growth initiative promotes energy-saving technologies, but it must keep the goal of a ‘sustainable future’ sufficiently empty to appease degrowth allies. Forming a stable hegemonic project is hence a precarious balancing act between the universal and particularist position. The demand for ‘sustainability’ gathers multiple, mutually divergent demands in a stable hegemonic formation by putting ‘sustainability’ forward as a universal demand, yet any concrete sustainability measure must fill in the void of the signifier’s meaning with particular content. Empty signifiers imperfectly condense heterogeneous aspirations into a single signifier able to mobilize political passions for a common hegemonic project (Islam, Holm, and Karjalainen 2017, 6; Szkudlarek 2007, 239). This dynamic opposition of purported universality and disavowed particularity is also present in the EGTAI.

3. TAI as a political unifier of the EGTAI: on the value of emptiness

The High-Level-Expert-Group (HLEG) on Artificial Intelligence released the EGTAI (HLEG 2019a) and the supplementary *Policy and Investment Recommendations for Trustworthy AI* (HLEG 2019b) in 2019. Before we dive in, we lay out why we focus on this policy instead of newer legislation. We want to show that the value of formulating ethics standards is controversial but not entirely ‘useless’, as Munn claims. Critical

discourse analysis shows that narratives matter, even without legal enforcement. Current legislative initiatives, like the AI Act, shift their narrative focus from political coalition-building and articulating hegemonic regulatory goals to the practicalities of legal enforcement. This process intersects with preexisting legislation in the EU and member states, as well as practical issues regarding execution and procedure. An analysis of the political struggle for hegemony between ethics and industry can abstract from these issues. The EGTAI offers an undiluted perspective on this tension – without reference to other legal frameworks – because its main goal is exactly to harmonize these demands without yet getting into the details of juridical enforcement. Thus, we are not concerned with how to make regulations more effective. On that front, initiatives like the AI Act and the DSA are admittedly more important. But the EGTAI is the political forerunner to this legislative trend. It articulates the conditions of possibility for the EU’s AI governance strategy, while later legislation goes further to put these principles into action. At the moment of writing, however, the AI Act and the DSA are not yet legally binding, so their influence will have to become visible in the next years.

Returning to the EGTAI, the first element to clarify is how the EGTAI invokes the different subject positions involved. The guidelines posit the promotion of ‘TAI’ as the overarching goal and claim to thereby bring three constituencies, or ‘pillars’, together. The documents favor a ‘holistic and systemic approach, encompassing the trustworthiness of all actors and processes that are part of the system’s socio-technical context throughout its entire life cycle’ (HLEG 2019a, 5). This entails a harmonious coordination of lawfulness, ethics, and robustness, or, in other words, a consensus among legal experts, ethicists, and the data experts from AI industry. The composition of the HLEG reflects this outlook (HLEG 2019a, 39): it has 51 members, with some legal experts (e.g. Joanna Goodey from the Fundamental Rights Agency), some academic ethicists or civil society representatives independent from business (e.g. Mark Coeckelbergh from Vienna University or Fanny Hidvegi from Access Now), and others from AI industry (e.g. Francesca Rossi from IBM). These individuals are assumed to represent legal, ethical, and industry demands as they exist in society as a whole. The EGTAI (HLEG 2019a, 6) explicitly minimizes the pillar of lawfulness, which is already discussed in other documents, to focus on the relations between ethics and AI industry, which we will do as well. The goal of ‘TAI’ supposedly unites the three pillars, but critics worry about the vagueness of the ‘TAI’-framework. AI industry can read ‘trustworthiness’ as a call for robustness, while ethicists and legal experts can simultaneously imagine that the document puts forward the agenda of making AI development more ethical and lawful. The critics interpret this vagueness as a failure, but, taking Laclau and Mouffe as our vantage point reveals a different reading. Rather than interpreting the semantic vagueness of ‘TAI’ as a failure to consistently regulate AI development, we should understand ‘TAI’ politically as an attempt to build a hegemonic coalition. Enforceable regulation would become the task of later legislative proposals, but the EGTAI sets out the political ambition to harmonize the subject positions required to make this legislative project possible.

The political function of ‘TAI’ is to build a chain of equivalences around an empty signifier ‘trustworthiness’ and create a consensual framework for future AI development. The EGTAI emphasizes that there are multiple important constituencies and that ‘in practice [...] there may be tensions between these elements (e.g. at times the scope and

content of existing law might be out of step with ethical norms)’ (HLEG 2019a, 5). Advancing a common goal like ‘TAI’ unifies these unstable elements in a political coalition. Vagueness is not a disadvantage but a necessity. Politically, ‘TAI’ enacts an outstanding inclusive achievement, given the diversity of the parties involved. Concepts kept open or empty are not meaningless but unite different voices into one canon. In the context of AI development, ‘TAI’ creates a negotiating space between divergent social demands that appears to be built on consensus while balancing various viewpoints simultaneously. Even if three distinct pillars are tied together, their binding glue is an agreement on using ‘TAI’ as a common negotiating platform. As the EGTAI specifies (HLEG 2019a, 24), the document only provides general guidelines for the implementation of TAI, but the concrete operationalization of these norms in particular cases of AI development cannot be determined in advance. The document leaves these debates to the stakeholders relevant to those particular cases, yet provides them with a framework for negotiating divergent demands. From Laclau and Mouffe’s perspective, the EGTAI hence produces an effective political coalition *through* empty signifiers, not despite of them. For unity to be kept intact, all stakeholders must agree on a universally applicable framework, which cannot be too specific and yet must be specific enough to be workable.

Nonetheless, empty signifiers do not produce a chain of equivalences among equals:

There is the possibility that one difference [i.e., one constituent subject different from other subjects in a chain of equivalences], without ceasing to be a *particular* difference, assumes the representation of an incommensurable totality. In that way, its body is split between the particularity which it still is and the more universal signification of which it is the bearer. This operation of taking up, by a particularity, of an incommensurable universal signification is what I have called *hegemony*. (Laclau 2005, 70, emphasis in original)

Within any political coalition, there will always be one subject hegemonic. This particular subject presents its own viewpoint as the universal perspective for the entire coalition. There is one constituent subject in the EGTAI whose interpretation of ‘TAI’ is hegemonic vis-à-vis the others. This hegemonic subject is continuously torn between voicing its own demands and keeping the meaning of ‘TAI’ empty enough to accommodate for alternative perspectives. In what follows, we argue that the EGTAI implicitly attributes hegemony to AI industry in three instances: the confusing chronology between demands from industry and ethics, the limited recognition of fundamental conflicts between industrial development and ethical imperatives, and the tendency to grant AI industry a stronger negotiating position in the assessment framework.

4. The hegemony of AI industry in the EGTAI

4.1. Ethics as a ‘fire extinguisher’ behind AI innovation

On first reading, the EGTAI seemingly privileges ethical demands over industry interests. It formulates *ethics guidelines* and its first chapter on ‘Foundations of Trustworthy AI’ puts AI ethics forward as its dominant source of inspiration (HLEG 2019a). A critical discourse analysis, however, suggests a different order of priority. The EGTAI explains the rationale for having ethics guidelines as follows:

It is therefore imperative that we understand how to best support AI development, deployment and use to ensure that everyone can thrive in an AI-based world, and to build a better future while at the same time being globally competitive. As with any powerful technology, the use of AI systems in our society raises several ethical challenges, for instance relating to their impact on people and society, decision-making capabilities and safety. If we are increasingly going to use the assistance of or delegate decisions to AI systems, we need to make sure these systems are fair in their impact on people's lives, that they are in line with values that should not be compromised and able to act accordingly, and that suitable accountability processes can ensure this. (HLEG 2019a, 9)

The goal of the EGTAI is 'to best *support* AI development, deployment and use', which suggests that AI industry holds the initiative with ethics as a support. The introduction to the *Policy and Investment Recommendations* contextualizes this claim: the US and China are already making significant advances in AI development, while the EU is 'in the exceptional position to put tailored policy and investment measures in place that can enable it to seize the benefits and capture the value of AI, while minimizing and preventing its risks' (HLEG 2019b, 7). The goal is to 'enable responsible competitiveness' in the market for AI (HLEG 2019a, 5). The EU allegedly worries about the many investments American and Chinese AI industries are attracting (HLEG 2019b, 43) and wishes to carve out a unique selling point for itself with 'TAI'. 'We also want producers of AI systems to get a *competitive advantage* by embedding Trustworthy AI in their products and services' (HLEG 2019a, 4, our emphasis). 'Trustworthiness' might hence be an ethical value that serves an economic purpose to distinguish European AI industry from its better-funded competitors. AI innovation is presented as the ultimate end, while 'TAI' is a justificatory label. 'TAI' constitutes an instrument to and not the goal of AI development, which complicates the supposed foundational position of ethics.

The language used to describe the role of AI ethics is often reactive and chronologically secondary to AI innovation. Championing 'TAI' is, for example, a way to respond to scandals from the past (HLEG 2019a, 33–35). If the EGTAI observes that 'the use of AI systems in our society raises several ethical challenges', it presumes that AI innovation precedes ethical reflection. Industry is responsible for developing AI systems, while ethics deals with the 'impact on people and society'. Ethics operates as a fire extinguisher: it minimizes collateral damage but is not a proactive agent questioning problematic innovations before they are implemented (Boenig-Liptsin 2022). If the EU renders 'TAI' an instrumental value in pursuit of competitive advantage, there is an incentive to *first* invest in AI innovation and only *secondly* assess the ethical impacts. The EGTAI lists a few fundamental rights that should limit the scope of AI development – respect for human dignity, freedom of the individual, respect for democracy, justice and the rule of law, equality, nondiscrimination and solitary, and, lastly, citizen's rights (HLEG 2019a, 11) – but it posits these rights as parameters against which the *effects* of AI innovation must be measured. In chronological terms, AI development hence comes first and, whenever it sparks a fire, AI ethics extinguishes it. Even when the EGTAI gives weight to ethics during the development phase, it is mainly in an advisory role, not as a method to directly intervene in or command the implementation process of AI, as when the EGTAI only recommends (without imposing) values-by-design methods (HLEG 2019a, 21).

4.2. The limited scope of EGTAI's tensions

The EGTAI provides space for disagreement and tensions, yet it limits the scope of these tensions in favor of AI industry's hegemony. Disagreement is allowed, but only within certain parameters. The EGTAI mentions that 'these guidelines are intended to foster responsible and sustainable AI innovation in Europe' (HLEG 2019a, 5), which implies that the imagined future is already assumed to entail more AI innovation. Fundamental questions whether to pursue AI-projects at all are *a priori* off the table. Ubiquitous AI is portrayed as the inevitable future. As Brevini (2021) writes, the EGTAI subscribes to a tech solutionism that represents AI innovation as a sufficient answer to humanity's most urgent problems. According to the EGTAI, it is beyond reasonable doubt that 'AI systems will continue to impact society and citizens in ways that we cannot yet imagine' (HLEG 2019a, 35). The only option is 'to build AI systems that are worthy of trust' (HLEG 2019a, 35). AI industry's interest in expanding the reach of AI in Europeans' everyday lives in search for profits is, however, put beyond question.

One passage where this ranking of priorities is particularly clear is the *Policy and Investment Recommendations*' section on AI and labor relations. The section speaks of 'a substantial reskilling need for every second employee within the next four years' and grants governments 'a key role in helping people anticipate, adapt, upskill, retrain and take advantage of the opportunities presented by new AI-related activities' (HLEG 2019b, 35). The document imagines the future of work as workers adapting to the pace of AI innovation. AI industry can pursue business opportunities according to its own standards and prospects, while ethics worries about the impact of this pursuit on workers, which it resolves by urging governments to prepare workers for an AI-dominated future. Workers' potential opposition to the algorithmic management of their labor is left un(der)imagined. Historically, the introduction of new technologies in the sphere of work has often been a contentious affair, leading to subsumption struggles between management and workers (Cant 2019; Mueller 2021; Jimenez Gonzalez 2022; Heemsbergen et al. 2022). The EGTAI claims that 'Europe needs to define what normative vision of an AI-immersed future it wants to realise' (HLEG 2019a, 9), yet it presupposes that the future has to be immersed in AI.

Admittedly, the EGTAI is not blind to the possibility of social conflict. It acknowledges the existence of tensions between the lawful, ethical, and robust pillars of TAI (HLEG 2019a, 5), but in the main and only section about these tensions, only conflicts *between ethical principles* are discussed (HLEG 2019a, 13). The EGTAI warns that the ethical principles mentioned in chapter I – Respect for human autonomy, Prevention of harm, Fairness, Explicability – create tensions. This seems considerate but overlooks tensions between industry interests and ethical demands. Conflict does not reside exclusively on the ethical level, given the diverse political coalition formed around 'TAI' as an empty signifier. The main area of political antagonism should not be located *within* the ethical pillar but *between* the different pillars. If we look at concerns with digital capitalism endangering AI ethics (Dixon-Román and Parisi 2020) or issues with algorithmic bias (Benjamin 2019; Noble 2018), then we mostly see the collision of ethics and industry's profit motive. When, for instance, hiring algorithms, like the one Amazon employed to sift through the CVs of managerial applicants, turn out to be biased against female applicants, the tension is not between different ethical values but between the

ethical value of fairness and the economic interest in cost-efficient CV-screening (Davies 2018; Criado-Perez 2019). Even if the EGTAI mentions that human oversight is essential when implementing AI (HLEG 2019a, 16), this does not mean that ethical standards will be prioritized by these human overseers. The real conflict resides on the deep disagreement between ethical and economic priorities.

4.3. *Unequal powers of assessment*

The EGTAI's critics make a fair point about the voluntariness of the EGTAI's assessment tools, which are presented as voluntary self-assessments rather than hard obligations (HLEG 2019a, 5). There is a significant risk that, if businesses are allowed to self-assess their conduct, AI industry will keep pushing through its own goals against ethical concerns (Bartoletti 2020). However, Smuha (2019, 101) is also right to insist that the HLEG's mandate from the European Commission did not grant them the power to formulate binding regulations, a task for subsequent legislative efforts. But the EGTAI still reveals some imbalances in the distribution of powers of assessment. Firstly, the EGTAI represents the nature of the disagreements that would come up during assessments in a very particular light. Ethical tensions should allegedly be resolved using 'methods of accountable deliberation [and] reasoned, evidence-based reflection' (HLEG 2019a, 13) or the 'moral universalism' of human rights (Smuha 2019, 103). If we read the EGTAI as an unstable coalition of divergent demands, such faith in deliberative reason or human rights misunderstands how deep disagreements work. If 'TAI' is an empty signifier fostering a chain of equivalences between essentially different perspectives, then tensions cannot be resolved through a mere appeal to reason or goodwill. If AI ethics and industry would directly deliberate on how to fill in and implement 'TAI', they would confront deep disagreements about what counts as 'evidence-based reflection' or 'accountable deliberation'. 'Human rights' can be interpreted very differently, so there is no guarantee a uniform and universal application of the 'Human rights'-paradigm exists (Mouffe 2005, 103). Ethical and economic outlooks on AI development and 'TAI' are *fundamentally* different, and the stability of their coalition depends on 'TAI' remaining somewhat vague and polysemic. Briefly zooming in on the assumption of one uniform rationality to adjudicate tensions, most computational approaches to AI define reason in terms of numbers and graphs, while ethical deliberation needs analogue conversation with points that cannot be 'proven' by numbers. Incommensurable conceptions of rationality are assumed to debate on a plain-level field, but this is not necessarily the case (Paul 2022, 502; Stamboliev 2023, 209). Mouffe (2005, 1) suggests that thinking about rationality as an immediately transparent and uniform method for consensus-formation demonstrates a post-political imaginary that disavows the nature of social conflict and deep disagreement. One acknowledges the possibility of conflict, yet simultaneously reduces it to a superficial obstacle easily overcome with a non-conflictual common understanding of rationality or human rights. However, if ethics and industry work with fundamentally different notions of rationality, there is not a uniform, non-contestable baseline for rational deliberation either. Political alignment occurs through hegemony rather than rational deliberation.

Mouffe herself already warns for the EU's tendency to devolve too much power to managerial technocracies while speaking the discourse of democratic control and

rational deliberation. One should not forget that the European Union itself is a hegemonic subject claiming to represent the ‘European people’ as a homogenous whole with shared fundamental values and norms. Mouffe (2013, 55) has thereby noted that the European Union tends to suppress internal social conflict in the name of these supposedly shared values. The fact that the European Commission chooses to appoint a group of experts, HLEG, rather than instituting a public debate among the European peoples evinces a post-political agenda. This expert group subsequently identifies itself as ‘we, as European citizens’ (HLEG 2019a, 4). Even if acknowledged as experts in their field, a particular group with a particular perspective then stands in for the whole European citizenry. Yet the HLEG does not only consist of legal scholars and ethicists, but also business representatives from European companies like Airbus and Zalando, and even non-European tech companies like Google and IBM (HLEG 2019a, 39). EU-discourse about ‘European citizens’ threatens to obscure the uneven representation of citizen interests in institutions like the HLEG.

Also, on the level of assessment procedures, discussed in chapter III on ‘assessing Trustworthy AI’, this implicit unevenness of power is apparent. That chapter delineates the assessment procedures for TAI, but regularly restricts its audience to industry players. Though the section briefly mentions the option of involving outside stakeholders (HLEG 2019a, 25), it also restricts that scope by recommending that ‘involving all stakeholders *in* a company, organization or institution fosters the acceptance and the relevance of the introduction of any new process’ (HLEG 2019a, 25, our emphasis). Ethical assessment seems a mostly in-house procedure conducted by AI industry itself. Subsequently, the EGTAI lists only intra-organizational stakeholders of AI development as relevant partners, like management, HR, or Quality Assurance departments. The assessment list’s formulation thereby hides a prioritization of who can produce rational evidence in this debate and who cannot. The main source of criticism for AI innovation allegedly comes from inside the corporations whose business model depends on expanding AI innovation. The subsequent pilot-version of the TAI assessment list (HLEG 2019a, 26–31) addresses an imagined ‘you’ that can only be interpreted as an upper-level intra-organizational stakeholder, which substantiates Reinhardt’s concern (Reinhardt 2022) that AI ethics principles often only address corporate agents. The list contains questions like ‘Did you carry out a fundamental rights impact assessment [?]’, ‘Did you consider the task allocation between the AI system and humans for meaningful interactions and appropriate human oversight and control?’, or ‘Did you assess the type and scope of data in your data sets?’. These are not meaningful question to ask an ordinary end-user of AI systems nor even a low-level employee of the presumed organizations. The latter would not conceive of data as ‘their data sets’. The imagined addressee is business owners and managers of AI-driven companies. The text ultimately attributes powers of assessment to a ‘you’ identified with AI industry.

For a document that claims to speak for ‘we, as European citizens’ (HLEG 2019a, 4) and champions ‘respect for democracy’ and ‘citizen’s rights’ (HLEG 2019a, 11), this distribution of power in assessment negotiations is noteworthy. Although the document *claims* that ‘all interested stakeholders can sign up to pilot the assessment list’ (HLEG 2019a, 24), the list itself is written for those with the power to make investment decisions and establish company policies. The ethical perspective would presumably only come from intra-organizational company departments like Compliance/SCR and Quality

Assurance, but these roles are technically guided and there could be conflicts of interest in arguing for companies to self-regulate and self-assess their own business initiatives (Bietti 2022; Perrigo 2022). Furthermore, it would be unrealistic to expect company employees to fundamentally challenge the business model of their employers. Ethical voices should sometimes be critical examiners of the AI life cycle rather than mere fire extinguishers. However, the EGTAI's portrayal of assessment procedures nudges conflicts between ethics and industry goals into a framework that favors industry over ethics, which hinders the latter's capacity to command a critical voice.

4.4. The hegemony of AI industry

We have highlighted the discursive elements in the EGTAI that suggest a hegemony of AI industry over ethics in the promotion of TAI. The diminished role of ethics as a fire extinguisher, the side-lining of conflicts between ethics and industry, and the power imbalances in assessing TAI suggest that industry directs the interpretation and implementation of TAI. However, we do not claim that ethics has no authority in the EGTAI or that it is mere useless veneer. A hegemony of industry does not imply that industry does not care about ethics or consistently silences ethical criticisms. Furthermore, this is a partial reading, focused merely on the tension between ethical demands and industry in one document. The EGTAI does not exhaustively settle the EU's position on AI regulation, which would require a broader study of all EU communications and legislative efforts. The most important message is that the TAI-framework unifies the EGTAI's discourse by simulating universal agreement on 'trustworthiness' yet simultaneously puts forward the particularist perspective of AI industry as the common-sense interpretation of 'TAI'. Ethics is thereby reduced to the role of the supporting act in the political coalition, rather than the foundational force, as the EGTAI seems to suggest.

5. Discussion: consequences of the hegemony of AI industry for the EGTAI

As the political basis for the EU's policy-framework concerning AI development, the EGTAI constructs a chain of equivalences of AI industry and ethics under industry's hegemony. It uses 'TAI' as an empty signifier to link their disparate subjects together in a single political coalition. Yet this does not definitively settle the future of AI. Hegemonic formations are discursively constructed entities open to contestation and change. The EGTAI articulates a fragile power-balance between different subjects, but that balance could still tip into different directions. Hence, the question remains what the future holds for AI ethics in the EU. Three scenarios are possible: a continuation of the status quo, assimilation, and contestation.

5.1. Status Quo

In the first scenario, *things stay as they are*. AI ethics is a distinct voice within the 'TAI'-framework yet stays under the hegemonic force of AI industry. AI governance would remain a fragile balancing act between economic and ethical demands, with the latter in a structurally subordinate yet not entirely powerless position. This contentious relation is today, for example, illustrated in the AI Act (Laux, Wachter, and Mittelstadt 2023). Laux

et al. observe unresolved tensions between ethics and industry, beginning with the EGTAI and continuing into the AI Act. The latter seems confronted with the impossible tasks of ‘engineering trust’ (16) and of managing the conflicting interests of opposing stakeholders without ethical trade-offs. Trustworthiness still seems the ethical compass for AI policy-making, yet it is caught up in an endless negotiation of divergent principles, like how to agree upon common standards for assessing risk. This weakens the guiding position of ethics yet keeps a seat at the table for ethical concerns. The chain of equivalences around ‘TAI’ grants some bargaining power to ethics by making industry dependent on the approval of ethics. It assigns ethics the role of granting legitimacy to AI innovation. That does not allow ethics much voice in directly steering the innovation agenda itself, which would be left to industry, but the latter would still need the ethics seal of approval. Ethically authorizing AI development in the name of ‘TAI’ would not be a mere empty gesture but gives ethics some bargaining power to force industry to negotiate for its agreement. The disadvantage of the status quo is, however, that this puts AI ethics in a paradoxical position, relegating the ethical pillar to a secondary (yet not insubstantial) role in a supposedly ethical policy-framework. That is particularly paradoxical given the document’s explicit aim to formulate *ethics* guidelines for TAI. The EGTAI does not empower ethics enough to *count* when it truly matters. What surfaces from the hegemonic formation of the EGTAI is that AI ethics is not given a priority in its *own* guidelines. Hegemonic formations are, however, fragile balancing acts that can tip over in two directions. Industry can either gain more power over ethics until eventually the latter becomes entirely integrated into industry’s business strategy, or ethics could contest its subordinate role and attempt to shift the balance of power in its favor. We propose a second and third potential scenario, called ‘assimilation’ or ‘contestation’. These will probably never correspond to either of these two extremes, yet they denote ideal-typical tendencies worth exploring to make sense of the possible futures of AI governance. If the status quo names a fragile balancing act between industry and ethical concerns, assimilation and contestation refer to the dissipation of this balancing exercise by either subsuming ethics entirely under industry interests or opting for open conflict.

5.2. Assimilation

The hegemony of industry over ethics could become so overbearing that ethics is entirely assimilated into industry’s demands. ‘Being ethical’ would then mean nothing more than enhancing business opportunities for AI industry. AI ethics becomes synonymous with supporting AI industry as an ‘ally’ and ethics would even lose the modest bargaining power it currently holds over industry. The EGTAI clearly does not condone such reductionism, but we highlight it as a potential risk for future AI governance. If one takes seriously the EGTAI’s claim that being ethical constitutes the EU’s competitive advantage on the global market for AI, then the promotion of AI ethics and the interests of AI innovation risk becoming fully equivalent. Ethics and industry would, in fact, become one and the same pillar, with ethics holding merely symbolic value to justify industry ends. Calling AI development ‘ethical’ would be mere decorative veneer without actually influencing the course of innovation in any significant way. Laclau calls this phenomenon ‘the unilateralization of the moment of subordination’ (Laclau 2005, 130): a hegemonic subject within a chain of equivalences weighs so heavily on the other, non-

hegemonic subjects that it absorbs them without remainder. This would no longer be a 'hegemonic formation', since the hegemonic subject has completely assimilated its partners. This possibility would substantiate the fears Floridi (2019) has voiced about phenomena like 'ethics shopping' and 'ethics blue washing'. AI industry would continue to use ethical labels but *only* insofar as they support the mission of expanding AI industry, not to encourage ethical policy-making in its own right. It mirrors a wider issue, namely, that ethics and industry are often entangled in terms of funding and agenda that they become codependent (Sætra, Coeckelbergh, and Danaher 2021). The 'TAI'-framework would then concern itself less with fostering ethically salient TAI than with deploying the language of ethics in whatever way befits the promotion of AI industry. Ethics would constitute a mere means to further business interests.

5.3. Contestation

An alternative is for AI ethics to contest its subordinate position in the hegemonic formation established by the EGTAI and advance its own position *openly*. This would mean that ethics would reject the current privilege of industry interests and dismiss the call for a consensual policy-framework. By contesting the rules of the game favoring industry, AI ethics would take a more explicitly critical and antagonistic attitude toward AI industry. According to Laclau, subjects previously mobilized in a particular chain of equivalences can defect and try to re-signify an empty signifier for rival political aims. Rather than accepting the framework of 'TAI' as established under the hegemony of industry, ethics could formulate its own, independent signifier of 'TAI' to rival industry's interpretation of the term. In those scenarios, Laclau (2005, 131) argues, empty signifiers become 'floating signifiers', i.e., signifiers of which the meaning floats between different, competing hegemonic projects. Their meaning becomes a stake in the struggle for hegemony. The collective identity of the coalition is thereby put into question and the collective falls apart into opposing camps that antagonistically attempt to acquire hegemony. This would resolve the paradoxical irony of the EGTAI in its current form insofar as ethics could claim its own voice in determining the *ethical* guidelines for AI development. Rather than bargaining for ethical reforms with industry under the latter's hegemonic leadership, the option of contestation encourages professional ethicists in academia or civil society to pursue its own vision of ethical AI innovation under its own hegemonic leadership.

Concluding our reading of the EGTAI, we support the option of contestation as this entails that AI ethics can *still* radicalize its vocation as critic of AI industry's agenda and re-signify 'TAI' as a framework that provides counterweight to industry interests. This would not only benefit AI ethics in advancing its own demands on its own terms but would also help to somewhat democratize policy-making by reinstating it as an arena of struggle over meaning (Mouffe 2005, 33; Rear and Jones 2013, 376). By shifting the focus of the EGTAI from 'common goals and aspirations' and deliberative consensus to hegemonic struggle with clearly distinct alternative visions for the future of AI development, policy-making could be opened up for more than one potential future. The meaning of 'trustworthiness' would become the site of direct social conflict rather than an unstable alliance between ethics and industry. Whereas industry could mainly be interested in 'TAI' in the sense of robustness, ethics could emphatically push for ethical 'TAI' even at the expense of

robustness. The two pillars would no longer have to be conceived as mutually compatible. The ultimate meaning of ‘trustworthiness’ would be a site of deep disagreement irresolvable through deliberative, evidence-based discussion. Rather than influencing AI development through its subordinate role within EGTAI as a necessary legitimation for the pursuits of AI industry, AI ethics could also take a more external standpoint and struggle against AI industry to hegemonise its *own* and *independent* understanding of ‘TAI’.

6. Conclusion

The EGTAI of the European Commission’s High-Level Expert Group discursively establishes a political unity around ‘TAI’ as an empty signifier, bringing together three camps: legal, ethical and industry representatives. Though each constituency would understand ‘TAI’ differently, the EGTAI succeeds at keeping the term empty *enough* to forge a chain of equivalences between these divergent subjects. However, this unity is not as equal as it might seem. AI industry occupies a hegemonic position vis-à-vis ethics in the chain of equivalences of the EGTAI. Though the document presents itself as emphatically a list of *ethics* guidelines for AI development, critical discourse analysis of the text reveals that the EGTAI (1) reduces AI ethics to the supporting role of a fire extinguisher subservient to AI industry’s projects, (2) limits the scope within which AI ethics can viably contest industry’s position, and (3) grants unequal powers of assessment to industry agents in the development and implementation of AI technology. Though the EGTAI presents itself as establishing a foundational ethics for AI development in the EU, it rather rearticulates the position of ethics vis-à-vis AI industry in a supportive role. Industry needs ethics’ authorization to pursue its own projects, which grants ethics some power to influence the course of AI development, but it still sets its own agenda and imposes it as hegemonic common sense. This puts AI ethics for a choice if it wishes to come to terms with the political conflict at the heart of AI development. Three paths are possible: AI ethics could continue to bargain for its demands from a subordinate position, as it does now. This would perpetuate the balancing act between industry and ethical demands within a hegemonic formation led by industry. However, AI industry could, secondly, extend its reach over ethical demands so pervasively that it subsumes ethics entirely and uses it as a decorative veneer for its own demands. In both scenarios, AI ethics confirms its subservient position to industry. But there is a third option: AI ethics could also contest the chain of equivalences built around industry’s hegemonic position. In this scenario, ‘TAI’ receives a contestatory value rather than acting as a signifier of reconciliation and consensus of seemingly opposing, yet uneven, pillars. ‘Trustworthiness’ is then an explicitly critical signifier meant to evoke a project of AI development that fundamentally rivals the demands of AI industry by supporting the ethical weight that trusting requires not only for AI but also for democracy.

Notes

1. While there is no general definition of AI, we take the following from the EGTAI: ‘Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or

unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal' (HLEG 2019a, 36).

2. Delineating the institutional politics of the EGTAI's creation moves beyond the scope of this article. See Smuha (2019) for an excellent overview from an insider perspective.

Acknowledgments

This paper was enabled by a visiting researcher fellowship from the Philosophy Department at Tilburg University, which was granted to Eugenia Stamboliev in 2022.

Disclosure statement

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

Funding

This work was supported by the Vienna Science and Technology Fund [10.47379/ICT20058].

Notes on contributors

Eugenia Stamboliev is a postdoctoral scholar and WWTF fellow at the University of Vienna, working on the philosophy and politics of 'AI' within democratic structures. She is affiliated with the Prague University of Economics and Business where she co-leads a research project on LLMs and democracy.

Tim Christiaens is assistant professor of economic ethics at Tilburg University in the Netherlands. He mainly works on the digitalization of work with a book on *Digital Working Lives* published with Rowman & Littlefield in 2022.

References

- Bartoletti, I. 2020. *An Artificial Revolution. On Power, Politics and AI*. London: The Indigo Press.
- Benjamin, R. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Medford: Polity.
- Bietti, E. 2022. "Self-Regulating Platforms and Antitrust Justice." *Texas Law Review* 101 (165): 165–202. <https://doi.org/10.2139/ssrn.4072084>.
- Boenig-Liptsin, M. 2022. "Aiming at the Good Life in the Datafied World: A Co-Productionist Framework of Ethics." *Big Data & Society* 9 (2): 205395172211397. <https://doi.org/10.1177/20539517221139782>.
- Brevini, B. 2021. "Creating the Technological Saviour: Discourses on AI in Europe and the Legitimation of Super Capitalism." In *AI for Everyone? Critical Perspectives*, edited by P. Verdegem, 145–160. University of Westminster Press.
- Brown, T. 2016. "Sustainability as Empty Signifier: Its Rise, Fall, and Radical Potential." *Antipode* 48 (1): 115–133. <https://doi.org/10.1111/anti.12164>.
- Cant, C. 2019. *Riding for Deliveroo: Resistance in the New Economy*. John Wiley & Sons.
- Coeckelbergh, M. 2020. *AI Ethics*. Cambridge: MIT Press.
- Coeckelbergh, M. 2021. *Green Leviathan or the Poetics of Political Liberty: Navigating Freedom in the Age of Climate Change and Artificial Intelligence*. New York: Routledge, An Imprint Of Taylor & Francis Group.

- Colpani, G. 2021. "Two Theories of Hegemony: Stuart Hall and Ernesto Laclau in Conversation." *Political Theory* 50 (2): 221–246. May. 009059172110193. <https://doi.org/10.1177/00905917211019392>.
- Colpani, G. 2022. "Two Theories of Hegemony: Stuart Hall and Ernesto Laclau in Conversation." *Political Theory* 50 (2): 221–246.
- Crawford, K. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.
- Criado-Perez, C. 2019. *Invisible Women. Data Bias in a World Designed for Men*. New York: Abrams Press.
- Davies, W. 2018. *Nervous States: How Feeling Took Over the World*. London: Jonathan Cape.
- Dixon-Román, E., and L. Parisi. 2020. "Data Capitalism and the Counter Futures of Ethics in Artificial Intelligence." *Communication and the Public* 5 (3–4): 116–121.
- Edelman, B., M. Luca, and D. Svirsky. 2017. "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment." *American Economic Journal: Applied Economics* 9 (2): 1–22. <https://doi.org/10.1257/app.20160213>.
- Eubanks, V. 2019. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.
- Floridi, L. 2019. "Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical." *Philosophy & Technology* 32 (2): 185–193. <https://doi.org/10.1007/s13347-019-00354-x>.
- Heemsbergen, L., E. Treré, and G. Pereira. 2022. "Introduction to Algorithmic Antagonisms: Resistance, Reconfiguration, and Renaissance for Computational Life." *Media International Australia* 183 (1): 3–15.
- HLEG (High Level Expert Group). 2019a. "Ethics Guidelines for Trustworthy AI." *Futurium - European Commission*. European Commission. <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>.
- HLEG (High Level Expert Group). 2019b. "Policy and Investment Recommendations for Trustworthy AI." <https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence>.
- Howarth, D. 2010. "Power, Discourse, and Policy: Articulating a Hegemony Approach to Critical Policy Studies." *Critical Policy Studies* 3 (3–4): 309–335. <https://doi.org/10.1080/19460171003619725>.
- Islam, G., M. Holm, and M. Karjalainen. 2017. "Sign of the Times: Workplace Mindfulness as an Empty Signifier." *Organization* 29 (1): 135050841774064. <https://doi.org/10.1177/1350508417740643>.
- Jimenez González, A. 2022. "Law, Code and Exploitation: How Corporations Regulate the Working Conditions of the Digital Proletariat." *Critical Sociology* 48 (2): 089692052110289. <https://doi.org/10.1177/08969205211028964>.
- Laclau, E. 1996. *Emancipation(s)*. London: Verso.
- Laclau, E. 2005. *On Populist Reason*. London: Verso.
- Laclau, E., and C. Mouffe. 2014. *Hegemony and Socialist Strategy: Towards a Radical Democratic Politics*. London: Verso.
- Laux, J., S. Wachter, and B. Mittelstadt. 2023. "Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and the Acceptability of Risk." *Regulation and Governance* Published online. <https://doi.org/10.2139/ssrn.4230294>
- Marchart, O. 2007. *Post-Foundational Political Thought Political Difference in Nancy, Lefort, Badiou and Laclau*. Edinburgh: Edinburgh University Press.
- Mittelstadt, B. 2019. "AI Ethics – Too Principled to Fail?" *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3391293>.
- Mouffe, C. 2005. *On the Political*. London: Routledge.
- Mouffe, C. 2013. *Agonistics: Thinking the World Politically*. London: Verso.
- Mouffe, C. 2018. *For a Left Populism*. London: Verso.
- Mueller, G. 2021. *Breaking Things at Work: The Luddites are Right About Why You Hate Your Job*. Verso Books.

- Munn, L. 2022. "The Uselessness of AI Ethics." *AI and Ethics* 3 (3, August): 869–877. <https://doi.org/10.1007/s43681-022-00209-w>.
- Noble, S. U. 2018. *Algorithms of Oppression. How Search Engines Reinforce Racism*. New York: New York University Press.
- O’Neil, C. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. London: Penguin Books.
- Paul, R. 2022. "Can Critical Policy Studies Outsmart AI? Research Agenda on Artificial Intelligence Technologies and Public Policy." *Critical Policy Studies* 16 (4): 497–509. <https://doi.org/10.1080/19460171.2022.2123018>.
- Perrigo, B. 2022. "Why Timnit Gebru Isn’t Waiting for Big Tech to Fix Ai’s Problems." *Time*. January 18, 2022. <https://time.com/6132399/timnit-gebru-ai-google/>.
- Rear, D., and A. Jones. 2013. "Discursive Struggle and Contested Signifiers in the Arenas of Education Policy and Work Skills in Japan." *Critical Policy Studies* 7 (4): 375–394. <https://doi.org/10.1080/19460171.2013.843469>.
- Reinhardt, K. 2022. "Trust and Trustworthiness in AI Ethics." *AI and Ethics* 3 (3, September): 735–744. <https://doi.org/10.1007/s43681-022-00200-5>.
- Ryan, M. 2020. "In AI We Trust: Ethics, Artificial Intelligence, and Reliability." *Science and Engineering Ethics* 26 (5): 2749–2767. <https://doi.org/10.1007/s11948-020-00228-y>.
- Sætra, H. S., M. Coeckelbergh, and J. Danaher. 2021. "The AI Ethicist’s Dilemma: Fighting Big Tech by Supporting Big Tech." *AI and Ethics* 2 (1, December): 15–27. <https://doi.org/10.1007/s43681-021-00123-7>.
- Smuha, N. A. 2019. "The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence." *Computer Law Review International* 20 (4): 97–106. <https://doi.org/10.9785/cri-2019-200402>.
- Stamboliev, E. 2023. "Proposing a Postcritical AI Literacy." *Media Theory* 7 (1): 201–232.
- Sunstein, C. 2017. *#republic: Divided Democracy in the Age of Social Media*. Princeton, New Jersey: Princeton University Press.
- Swyngedouw, E. 2010. "Apocalypse Forever?" *Theory, Culture & Society* 27 (2–3): 213–232. <https://doi.org/10.1177/0263276409358728>.
- Szkudlarek, T. 2007. "Empty Signifiers, Education and Politics." *Studies in Philosophy and Education* 26 (3): 237–252. <https://doi.org/10.1007/s11217-007-9033-7>.