AI & SOCIETY (2021) 36:597–608
https://doi.org/10.1007/s00146-020-01073-0

OPEN FORUM

An emerging AI mainstream: deepening our comparisons of AI frameworks through rhetorical analysis

Epifanio Torres[1] · **Will Penman**[2]

Received: 30 July 2020 / Accepted: 27 August 2020 / Published online: 22 October 2020

Abstract
Comparing frameworks for AI development allows us to see trends and reflect on how we are conceptualizing, interacting with, and imagining futures for AI. Recent scholarship comparing a range of AI frameworks has often focused methodologically on consensus, which has led to problems in evaluating potentially ambiguous values. We contribute to this scholarship using a rhetorical perspective attuned to how frameworks shape people's actions. This perspective allows us to develop the concept of an "AI mainstream" through an analysis of five of the highest-profile frameworks, including Asimov's Three Laws. We identify four features of this emerging AI mainstream shared by most/all of the frameworks: human-centered design focus, abstraction-oriented ethical reasoning, privileged authorship, and ahistorical regulatory justifications. Notably, each of these features permeates each framework, rather than being limited to a single principle. We then evaluate these shared features and offer scholarly alternatives to complement and improve them.

Keywords Human-centered · Abstraction · Authorship · History

1 Introduction

With the ongoing development of artificial intelligence (AI) technologies, new frameworks for regulating AI are increasingly being developed. By one count, 36 frameworks were issued in 2018 alone, making 84 frameworks total since 2011 (Jobin et al. 2019). As a result, scholarship that does some initial sensemaking is helpful. Based on consensus among six high-profile frameworks, Floridi and Cowls (2019) identify five "core principles" (n.p.). Similarly, Jobin et al. (2019) use a consensus-driven analysis to identify eleven "overarching" ethical values/principles, with five principles referenced in more than half of the frameworks. And for Hagendorff (2020), it is consensus in 22 recent frameworks that helps uncover AI goals to critique. In the process of identifying consensus among frameworks, these scholars have also found it appropriate to make various critiques. Hagendorff puts it most strongly by asserting, "AI ethics is failing in many cases" (p. 113).

Yet with such a strong methodological focus on consensus, existing scholarship has little way to differentiate between positive and negative convergence. Jobin et al. (2019), for instance, critique several areas in which the frameworks converge, including that many of them come from a similar geographic area and that many of them do not emphasize sustainability. But surprisingly, these critiques (and other critiques based on convergence) are relegated to a discussion section and actually represented in the abstract as divergence! Floridi and Cowls (2019) encounter a different problem from emphasizing consensus. They argue that we should include China in our ethical discussions but also argue that the consensus they have already identified is sufficient (p. 9). Meanwhile, Hagendorff (2020) devotes almost his entire paper to insightful critiques of how AI frameworks are similar, but ultimately, only theorizes his concerns as ethical "omissions," which implicitly minimizes his own critiques of the AI frameworks' content. All of these share a consensus-driven focus that leads to other conceptual problems.

Rooted in a rhetorical approach, we take a stance that acknowledges the ambiguity of consensus itself. While consensus can sometimes represent a group's resolve,

✉ Epifanio Torres
epifanio@princeton.edu

Will Penman
wpenman@princeton.edu

[1] Department of Computer Science, Princeton University, 35 Olden St, Princeton, NJ 08540, USA

[2] Princeton Writing Program, New South 308, Princeton, NJ 08540, USA

cooperation, and mutual empathy, consensus can also indicate a group's recalcitrance, narrowness, and selfishness. Thus, we seek to identify a range of similarities among AI frameworks while putting ourselves in a position to critically evaluate the social trends they reflect and develop. To begin identifying similarities, we focus on high-profile and comprehensive frameworks. High-profile frameworks, as Floridi and Cowls (2019) observe, can more visibly influence society, which lends our analysis some face validity; likewise, comprehensive frameworks present a complete set of guidelines which allows for a thorough analysis. On this basis, we have selected the following five high-profile and comprehensive frameworks:

1. "Ethically Aligned Design" by the Institute of Electrical and Electronics Engineers (IEEE). The IEEE document has developed in stages by committee, with the First Version released in 2019. At 294 pdf pages, the framework is formidable. Early chapters elaborate on "ethically aligned design," and then eight chapters discuss various topics ranging from Well-being to Law (IEEE 2016).
2. "Ethics Guidelines for Trustworthy AI" by the High-Level Expert Group on Artificial Intelligence (AI HLEG), a group of 52 (AI HLEG 2019, p4) scholars and European stakeholders convened to advise the European Commission. AI HLEG published a draft framework in late 2018 and published a revised version in mid-2019 after receiving extensive public feedback. The bulk of the 41-page report develops a framework centered on "trustworthiness" (AI HLEG 2019). At the end of the document, various opportunities and concerns about AI development are examined.
3. "AI Principles" and "Responsible AI Practices" by Google. The Principles originated as a mid-2018 blog post from CEO Sundar Pichai (Pichai 2018), and the Practices were offered six months later as a complement, along with a review structure for applying the principles (Walker 2018). The Principles provide a numbered list of seven objectives (leading with "Be socially beneficial"), each with a short explanation. They are followed by Google's commitment to four AI areas the company will not pursue. Currently housed in the "Responsibilities" tab of the Google AI subdomain, the web pages are occasionally updated and are, therefore, not as textually stabilized as the other frameworks.
4. "The Beijing AI Principles" by the Beijing Academy of Artificial Intelligence (BAAI) and Chinese scholars/industry leaders. In mid-2019, the BAAI published their framework as a brief blog post in English on their otherwise Chinese-language website. The post briefly frames the document as a call to develop "beneficial AI for humankind and nature," then consists of 18 bulleted maxims grouped under "Research and Development," "Use," and "Development" (BAAI 2019). Unlike the other frameworks, the BAAI framework has been interpreted in the media as having political support from the government (Knight 2019; Chia 2019). BAAI has significant Chinese-language web presence.
5. Finally, "The Laws of Robotics" by Isaac Asimov. Asimov's framework was first published in a short story in 1942 and later developed in the 1950 collection of short stories, *I, Robot*. It consists of three "laws" that a lifelike robotic system must operate with: not to injure humans (First Law); to obey humans (Second Law); and to preserve itself (Third Law) (Asimov 2004). Asimov later added a Zeroth Law, intended to supersede the First Law: not to injure humanity. (We also incorporate Asimov's narratives along with the Three Laws themselves, because they serve as the practical manifestation of the principles.) Asimov's framework is unique for being established and tested in a fictional world, before deep learning AI techniques were even imagined. Among other markers of being high profile, it has been the subject of a blockbuster movie in 2004 (*I, Robot*, grossing $346 million worldwide [Proyas 2004]) and has been extended and updated by legal scholars today (Balkin 2017).

Each year, more frameworks are being developed, but these five represent a thorough and diverse picture of today's highest-profile frameworks. To give full weight to the varying lengths and complexities of the frameworks, we discuss them from longest to shortest: IEEE, AI HLEG, Google, BAAI, and Asimov. Throughout this paper, we use "AI" as a cover term that uniformly addresses the subjects of the frameworks.

By looking at several high-profile frameworks together, we are able to document a kind of "AI mainstream." The use of "mainstream" describes ideas' power (rather than rightness), hints at alternatives, and consolidates existing scholarly insights that deal with either the positive or the negative aspects of consensus by including both. In theorizing an emerging AI mainstream, we find it useful to draw on the field of rhetoric, because it focuses on context and the social function of discursive products like AI frameworks. From a rhetorical perspective, "mainstream" is an articulation of the concepts of "*doxa*" and "cultural logics." Correspondingly, the ambiguity expressed by "AI mainstream" gives insight into our responses as a society, namely for us to develop contextual wisdom (i.e., *phronesis*). This can produce a helpful skepticism towards consensus that allows us to be mindful of the ideas that we are incorporating into the mainstream and those that we are not. As we describe it, an emerging AI mainstream has at least four features: a human-centered design focus, abstraction-oriented ethical

reasoning, privileged authorship, and ahistorical regulatory justifications. These features echo, develop, and qualify existing scholarly discussions.

Our analysis also allows us to develop the scope of the emerging AI mainstream. We make the counterintuitive choice to include Asimov's "Three Laws of Robotics" in our analysis, which is explicitly excluded from other comparisons of AI frameworks (Floridi and Cowls 2019; Hagendorff 2020; Jobin et al. 2019). For many commentators today, Asimov's fictional setting is a misleading test site for AI regulations: his focus on robots is too narrow for today's range of AI uses; the laws themselves are not able to be implemented at a technical level (much less with a hardwired "positronic brain"); and overall, Asimov's framework addresses issues related to robots' power that seem tangential to the issues of surveillance, military machines, and bias facing AI development today (Singer 2009; Dvorsky 2014; Anderson 2017). These objections would predict extreme dissimilarity from the other frameworks which are rooted in the "real world." In the midst of these objections, our choice to include Asimov follows from a rhetorical emphasis on flows of discourse and power. That is, popular culture is equally invested in—and democratically part of—technical and legislative directions. Through finding that Asimov equally participates in the emerging AI mainstream, we argue that additional scholarly attention should be given to other frameworks and narratives developed in fictional, imaginative settings. Correspondingly, this finding also suggests that the emerging AI mainstream is, for better or for worse, relatively robust.

From here, we first examine relevant rhetoric scholarship to introduce terms that guide our identification and discussion of features. Then, we examine four shared features among most/all of the frameworks: human-centered design focus, abstraction-oriented ethical reasoning, privileged authorship, and ahistorical justifications. Afterward, we critically return to each one in order to evaluate this emerging AI mainstream and identify alternatives. We generally support a human-centered design focus but argue that the other three features should be reconsidered as we engage with AI technologies into the future. We conclude by pointing to directions for future research in this emerging space.

2 Literature review

In the interdisciplinary spaces of investigating AI, scholars have found it useful to draw on disciplinary conceptions (e.g., Floridi and Cowls 2019 ground their comparison of AI frameworks in the discipline of bioethics). In this paper, we use tools from the field of rhetoric to help make sense of AI frameworks. Rhetoric has been traditionally focused on the art of persuading others; a rhetorician's skill is in flexibly adapting their appeals in context-specific ways to the audience and situation at hand. Rhetoric has been recently extended to include our natural propensity for being persuaded ourselves (Davis 2010) and to include related listening skills for thoughtfully hearing other perspectives (Ratcliffe 2005). That is, insofar as rhetoric is focused on analyzing specific texts (like AI frameworks), it is also tuned to how different ideas, beliefs, and actions flow back and forth among people, animals, and the built environment. This flow is particularly important for interpreting AI frameworks, because frameworks often explicitly seek to shape people's ideas, beliefs, and actions, and to extend that influence to the AI products that we create.

At an initial level, the rhetorical concepts of *doxa* or "cultural logics" give intuition for why we might expect an AI mainstream to be emerging. *Doxa* (from ancient Greek, adjectival form: doxastic) describes a culture's sometimes unconscious widely-held beliefs; these are important for rhetors to attend to because they are easy starting points for rhetors to connect with their audiences. In the case of AI, for instance, we might identify the idea that "AI is going to continue to be developed" as doxastic. That is, no framework that we have seen addresses whether we *should* continue developing AI in some form or another; for readers of these frameworks, it is doxastic—of course AI is going to continue to be developed—so our frameworks can start from there and take a guiding role. Because any society has *doxa* on a range of issues, it would not be surprising from a rhetorical perspective for us to have been developing doxastic commonsense around AI, too. In fact, because there are overlaps between genres and between large spheres of discourse (Schiappa 2012), we would expect that high-profile popular culture (perhaps even Asimov) might be informing—in a subtle way—the technical development of AI frameworks. Krista Ratcliffe (2005) draws this idea tighter by conceiving of these shared ways of approaching the world as "cultural logics." The plural "logics" indicates that *doxa* are not just conclusions, but whole logical patterns and ways of making sense.

Rhetorical theory furthers this by making us skeptical of consensus. In a classical sense, rhetors who seek to persuade people of something unpopular need ways to *dislodge* consensus; proponents of civil rights have long observed the difficulties of doing so and the injustices from people collectively stopping up their ears (Johnson, 2012). That is, consensus is not to be uniformly lauded or dismissed (Roberts-Miller 2007; Flower 2008), nor is it sufficient to develop deliberative forums on their own (Jobin et al. 2019). In a technological context, we see this applied by Siby K. George, who advocates for a skepticism towards ideological consensus. According to George, the Global South has faced mounting pressure from much of the Global North to conform to "techno-capitalism" for the sake of development. George encourages us to rethink our *doxa* so as not to stamp

out other possible socio-economic arrangements or dismiss the value of different perspectives within the Global South (George, 2017). Thus, identifying an AI mainstream often means identifying points of ethical ambiguity. "Mainstream" ethical reasoning for AI fundamentally describes ethical influence and flow and does not straightforwardly translate to evaluation. Rather, we must use responsive skills on a case-by-case basis (Ratcliffe 2005)—*phronesis*, or applied wisdom.

In addition to tuning us to the contours of an AI mainstream, rhetorical theory helps our inquiry methodologically. First, we analyze the frameworks in terms of their rhetorical appeals, or opportunities for "identification" (Burke 1966), finding a shared human-centered design. Second, we analyze the frameworks in terms of their reasoning strategies (or *topoi*), finding a mostly shared abstraction-oriented ethical reasoning. Third, we analyze the frameworks' ethos (which when explicitly shared has been called "interdependence" [Penman 2018]), finding a shared privileged authorship. Finally, we analyze each framework's justifications for its guidelines, or rhetorical "exigence," finding each framework's justifications to be relatively ahistorical.

3 Four shared features of an emerging AI mainstream

In this section, we identify four features that are shared among all (or most) of the five frameworks under our consideration. Although many of these have been identified by existing comparisons, we attend to how these features emerge in each framework and permeate throughout each one's composition. This extensive descriptive work sets us up to better evaluate and critique these features of the emerging AI mainstream in the following section.

3.1 Human-centered design focus

A prominent shared feature among the five frameworks is an overarching design focus on humans. Designing AI for humans means rejecting placing AI and humans on equal standing and rejecting treating AI systems as a kind of life-form that inherently needs to be preserved. We identify human-centered design through explicit anthropocentric terminology, observing that it permeates each framework rather than only emerging as an individual principle. We then identify how a human-centered design focus functions rhetorically through its implicit contrasts.

In the case of IEEE, a human-centered design focus emerges early in the document, and a strong sense of anthropocentrism is maintained throughout. For instance, IEEE uses "human rights" as an organizing principle, appearing in every section of the document (p. 13). Other guiding principles, like "human well-being," are both anthropocentric and used to establish humans "as a primary success criterion" for AI development (p. 11).

Much like IEEE, AI HLEG establishes a human-centered design focus early in its introduction by asserting that AI needs to be "*human-centric*" (p. 4, emphasis in original). AI HLEG's overarching concept of "trustworthy AI" directly means trustworthiness *to humans* (p. 4). Thus, the framework's human-centered design focus underlies the entire document and influences other related terms, like how AI systems should be "user-centric" for a range of *human* users (p. 18).

Similarly, Google's Practices document calls for "a human-centered design approach" (n.p.) and Google' Principles document lists several principles that derive their meaning by reference to humans, ranging from being "socially" beneficial to incorporating "privacy" considerations (n.p.). Additionally, Google's framework establishes a user-tool dynamic in which humans are superior and AI systems are "tools" that must be "accountable to people" (Principles, n.p.).

For BAAI, the concept of human-centered AI operates within a broader environmental understanding. This is seen in the first sentence, in which AI concerns "the future of the whole society, all humankind, and the environment" (BAAI). This environmental conscientiousness pervades the framework; AI that benefits "all humankind and the environment" (alternatively, "society and nature," n.p.) is used to guide recommendations for research and governance both now and in the future. Interestingly, BAAI explicitly includes "the external environment where the AI system deploys" (n.p.) in its discussion of safety.

Finally, in the case of Asimov's framework, both the Zeroth Law and First Law put human existence and well-being at the peak of the framework's ethical hierarchy. The Second Law goes a step further by establishing robots' subordination to humans. In the story "Runaround," this produces a kind of awkwardness when humans interact with social AI systems, which were built with "slave complexes" and which called humans "Master" (p. 29). And despite Asimov's extreme emphasis on robot subordination, the Third Law surprisingly makes concessions for robots' well-being.

We also identify the rhetorical functions of being "human-centered" through the frameworks' implicit contrasts. First, human-centered design focus affirms human autonomy in the midst of AI's integration into society. For instance, both Google and AI HLEG affirm that lethal autonomous weapons *are not* human-centered; as AI HLEG puts it, this is because "human control is almost entirely relinquished" (p. 34). Second, a human-centered design focus dissuades using AI to accumulate wealth, power, and other privileges at the expense of others. ("Human-centered" in this usage takes vulnerable people as the prototype of "human.") For

instance, in commenting on the principle "For humanity," BAAI explains that "AI should not be used to against [*sic*], utilize or harm human beings" (n.p.). IEEE also dissuades using narrow metrics for human welfare, advocating for "the full spectrum of well-being" including "psychological, social, economic fairness, and environmental factors" (p. 6). Finally, the frameworks' emphasis on being human-centered counters potential ontological flattening in which humans and AI are viewed as "the same." For instance, AI HLEG inveighs against AI systems that do not disclose themselves as such (p. 34). Likewise, this theme is of special public concern in Asimov's story "Evidence," in which a mayoral candidate's potential robotic identity causes public investigation and almost leads to corporate scandal and legal debacle.

Overall, we find that the frameworks overwhelmingly center AI design on humans. Explicit human-centered terms are used early and often to set humans as a primary success criterion for AI, and other values/principles often obtain their meaning by reference to humans. Human-centered design focus can include environmental conscientiousness. Moreover, designing AI to be human-centered is complicated by instances where humans and AI both act as social agents. Finally, human-centered design functions rhetorically to support human autonomy, equality, and uniqueness.

3.2 Abstraction-oriented ethical reasoning

A second point of analysis consists of understanding how recommendations for AI development interact with values that we want to achieve and maintain. To identify this relationship, we examine how the frameworks graphically and textually motivate their concrete recommendations. We find that almost all of the frameworks allow their most abstract values to guide and limit their most concrete recommendations—reflecting "abstraction-oriented" ethical reasoning. That is, most of these frameworks do not include a bidirectional flow in which practices inform ideals, too. Even more so than with human-centered design, abstraction-oriented ethical reasoning is an underlying way of *structuring* frameworks, not a discrete principle within any given framework.

Graphics can be a powerful form of representing document-level reasoning and structure, and two frameworks use graphics in a way that demonstrates abstraction-oriented reasoning. In IEEE, a one-directional flowchart shows how ideas move "*from* principles *to* practice" (p. 15, our emphasis). Three "pillars" reflect the abstract ideals of "universal human values," "political self-determination and data agency," and "technical dependability." In the chart, an arrow goes from these to the right towards "general principles," indicating how the principles are informed by the pillars. From the general principles, another right-ward arrow goes to each chapter, indicating that the chapters are informed by the principles. From each chapter, a final arrow goes to "issues" and "recommendations." Visually, then, practical actions ("recommendations") are the outcome of applying abstract principles. The figure layout and sequential arrows show a one-directional flow, in which values restrain, condition, and evaluate possible actions for dealing well with AI. Similarly, AI HLEG includes a directional flow chart that clearly visualizes the framework's abstraction-oriented ethical reasoning through its downward arrows. Four abstract ethical principles are placed at the top: "respect for human autonomy," "prevention of harm," "fairness," and "explicability (p. 8). From these, a downward arrow indicates seven key requirements for realizing trustworthy AI. The requirements, in turn, have another downward arrow to technical and non-technical methods for implementing them. Thus, taking action with regard to AI ("methods") is at the end of the chain; it is informed and regulated by abstract principles. Moreover, this abstraction-oriented reasoning structures the document itself, from Chapter 1 (principles), Chapter 2 (requirements), and Chapter 3 (methods).

Abstraction-oriented ethical reasoning also plays out textually in how every framework (except for BAAI) rationalizes its recommendations. As we would expect from IEEE's and AI HLEG's graphics, IEEE explicitly conceives of norms as hierarchical (p. 174–175). At a chapter level, five out of six issues on law directly draw on language from IEEE's general principles (e.g. "Issue 1: Well-being, Legal Systems, and A/IS" references the abstract value of "well-being" p. 214). Obtaining issues from principles instantiates an abstraction-oriented reasoning structure. AI HLEG also motivates concrete practices by reference to agreed-on abstract principles. For instance, a typical assessment item is a straightforward outcome of the framework's more abstract principles: "Did you establish mechanisms that facilitate the system's *auditability*, such as ensuring traceability and logging of the AI system's processes and outcomes?" (p. 31, emphasis ours). The item's appeal to auditability is legible to readers as important because it references the technical method of "Auditability," which itself is justified as a measure of the ethical principle of "Explicability," which is in turn "rooted in fundamental rights" (p. 11). Thus, the organizational mapping of AI HLEG is one-directional; the practices are prevented from meaningfully influencing or informing its guiding abstract principles.

Google's framework implicitly follows this pattern of prioritizing principles over practices. Chronologically, the Principles came first, and then were followed by the Practices. Additionally, each of Google's four topic pages in their Practices clearly maps to a different principle (e.g., "Safety" echoes the principle "Be built for safety"). And yet, unlike IEEE and AI HLEG's repeated appeals, these connections remain implicit because Google's Principles and Practices are located in separate webpages and are not hyperlinked to each other. In fact, the Practices are prioritized in the

website navigation, corporate leaders emphasize the interconnectedness of the two (with the Practices being "a complement" to the Principles [Walker 2018] and the Principles being "supported with" [Dean 2019] the Practices), and the Practices document even appeals to "general best practices for software systems" for its guidance rather than abstract principles. These aspects soften the hierarchical, abstraction-oriented effect identified in the other frameworks.

Asimov's laws explicitly rank principles, and each subsequent law from the First/Zeroth comes with an "as long as" qualifier. These "as long as" qualifiers set up an explicit hierarchy of principles for robot action. Moreover, the story "Liar" deals with the challenges of practically implementing the First Law's abstract decree. In the story, a mind-reading robot discovers that in practice, causing emotional harm is sometimes unavoidable. And yet, this practical discovery is unable to be incorporated into the robot's sense of the First Law, and the narrative resolves when the robot short-circuits and goes "insane" (n.p.). This story graphically illustrates how the abstract principles in the Three Laws influence a robot's actions, and never the other way around.

When compared to the structures of the other four selected frameworks, BAAI emerges as an outlier. The framework itself is grounded mostly in recommendations for actions that "should" (n.p.) be performed by humans and AI rather than being explicitly grounded in abstract principles.[1] Moreover, the document is organized in sections that reflect the "lifecycle" of AI (Research and Development, Use, and Governance), rather than separating principles from practices. As a result, BAAI least reflects abstraction-oriented ethical reasoning.

3.3 Privileged authorship

Our third point of analysis examines the frameworks' authorship. This is a kind of "metadata" that shows who has been given the power to create frameworks and regulations for AI. We first discuss how each framework recognizes the importance of incorporating diversity and feedback; then we examine the global standing and status of the authors.

Initially, the five frameworks seem to seek the inclusion of a range of perspectives. IEEE, AI HLEG, Google, and BAAI express a special concern for the marginalized. Additionally, IEEE received more than 500 comments during the drafting process and acknowledges that different stakeholders and committee members had disagreements over some of the particulars. Similarly, the AI HLEG received over 500 pages of feedback during the drafting process and signaled a polyvocal stance in their recommendations. Google also asked the general public to provide them with feedback on its principles. At a system level, Google introduced system-driven efforts like diversity lectures to employees and a "fairness" module for machine learning students, and promotes an ongoing practice of "be[ing] accountable to people" (Pichai 2018). BAAI makes one of its eighteen maxims about this very aspect: "Be Diverse and Inclusive" (n.p.). Even Asimov's narratives gesture towards the inclusion of marginalized groups despite being written in the 40s and 50s. For example, his narratives show the open acceptance of a robot religion that draws strong parallels with Islam and position an educated and pragmatic woman as the main protagonist of several stories.

With that being said, after analyzing the authorship of the selected frameworks, we have found that the authorship of each framework generally represents the voices of the privileged. Based on a visual assessment of the IEEE's supplementary document which lists the different authors and contributors of the framework, several people of color seemed to form part of its executive committee and aid in the development of the framework's different chapters. Nevertheless, most contributors with personal descriptions come from elite circles of society within academia, business, and law. This suggests that IEEE includes public feedback in a limited sense where the executives, scholars, and lawyers get the final say on what gets included in the document. Similarly, AI HLEG was convened to be a "high level expert group," mostly representing elite academic and business institutions. Google's framework was originally published in blog posts attributed to Google's CEO and SVP of Global Affairs. In today's web page format, authorship is hidden and attributed to the company overall ("Google is committed…"). Compared to the others, BAAI has a complex positioning in terms of authorship. Although the framework comes from an underrepresented region within English-language academia and business, the framework itself was "developed in collaboration with the most prominent and important technical organizations and tech companies working on AI in China" (Knight 2019), suggesting BAAI still reflects a notably elite authorship and decision-making process. Asimov also has a complex authorship, being the sole author of *I, Robot* and ultimately having "say" over the content of the Laws of Robotics, but presenting the narratives in *I, Robot* as an oral history by someone who takes her own perspective on events.

As seen above, each of the frameworks acknowledges, at least to some extent, the benefits of treating all people (including the minoritized and underprivileged) with fairness and respect. At the same time, in all five frameworks, people who are already powerful retain the ability to define right and wrong (i.e., encouraged and discouraged, permitted and prohibited) uses of AI.

[1] BAAI does use the word "principles" but is naming actions that must be taken collectively, not abstract values.

3.4 Ahistorical regulatory justification

Finally, we examine how each framework justifies its own contribution to AI development, particularly in relation to the past. Being historically aware grants us the vision to see beyond present social structures to understand the underlying historical trends that cause and inspire them. This helps us to wisely identify exigences as we seek to grapple with our past and negotiate our future. Our analysis of the five selected frameworks reveals that all of the frameworks recognize that AI has the potential for abuse; several of them weakly address this danger as ongoing; and none of them addresses a strong form of historical reckoning with their own involvement in oppression. Thus, all of them operate in a relatively "ahistorical" justificatory space.

The simplest form of incorporating history, which all of the frameworks adopt, is recognizing that emerging AI has the potential for abuse. IEEE makes a pointed recognition of how Western traditions have dominated ethics. In an effort to counteract that dominance, the document has some discussion of other ethical systems as well, such as Buddhism, Confucianism, and Ubuntu traditions (p. 49). Similarly, AI HLEG establishes a fundamental right to "equality, non-discrimination and solidarity—*including the rights of persons at risk of exclusion*" (p. 2, emphasis ours). This acknowledges the dangers of algorithmic discrimination produced due to harmful social biases. BAAI makes explicit references to "reducing possible discrimination and biases." Finally, the Three Laws themselves are represented as a legal and business concession made to reduce public fears of robotic development. However, a mere recognition for the potential for abuse (as seen in these frameworks) does not necessarily engage with the past.

A deeper form of justifying the frameworks' contribution, which only some of the frameworks adopt, is to recognize that these harms are ongoing because they come from historical precedent. IEEE acknowledges that AI could be "perpetuating" or even "exacerbating" historical bias in AI systems (p. 258). Likewise, AI HLEG takes up direct and sustained discussion about the potential dangers of the perpetuation of racial discrimination by AI systems. Similarly, Google acknowledges the need to be mindful of "problematic pre-existing biases" based on race, gender, and others (Practices 2018). BAAI weakly addresses historical patterns through addressing "those who would otherwise be easily neglected or underrepresented in AI applications." And Asimov does not recognize that today's AI risks may stem from historical patterns.

Finally, we find that none of the frameworks reckon with their own past involvement in the midst of their recommendations. Although IEEE guides developers to use "user-level" and "community-level" (p. 184) metrics of success for AI systems that embed values and to incorporate members of disadvantaged groups in these processes (p. 188), they do not justify these guidelines with reference to any ways that engineers and developers have historically failed to incorporate these values. In AI HLEG, "stakeholders" are nominally defined to include users of AI systems and even those impacted by AI systems (p. 37–38), but the implementation guidelines are oriented toward only those already in power, such as lawmakers, engineers, and executives. For Google, it is likely that corporate communication norms limit renouncing their participation in ongoing structural wrongs. Stylistically, Google's Practices document and Principles document do not have much self-involvement or express a recognition of Google's past involvement in technological inequity. In more direct ways than the AI HLEG, Google's framework still keeps the ultimate ethical decision-making power in the hands of the executives, since Google's committee still makes the final decisions (Walker 2018). BAAI fails to include its past involvement in discriminatory practices, which has been perceived as hypocritical given China's use of AI to discriminate against Uyghers (Knight 2019; Chia 2019). Finally, in Asimov's stories, the engineers are able to make small tweaks to weight each law, but the laws themselves are enforced and applied at a hardware level without human intervention.

4 Evaluating similarities

Our analysis of rhetorical elements of five high-profile frameworks has revealed an emerging AI mainstream with at least four features. One function of the term "mainstream," as we indicated in the introduction, is to effectively integrate evaluative work of these features, in order to allow for a more nuanced and inclusive discussion around AI and alternatives to mainstream trends. Thus, we re-examine each feature with an eye toward scholarship that engages and challenges the mainstream. Specifically, we support, with qualifications, a human-centered design focus for AI, and thereby focus our discussion on possibilities for continuing this. In contrast, we find significant limitations in the features of abstraction-oriented ethical reasoning, privileged authorship, and ahistorical justifications.

4.1 Possibilities for a human-centered design focus

The first shared feature that we identified was a "human-centered" design focus. Other scholars have also identified this similarity among frameworks (Floridi and Cowls 2019; Jobin et al. 2019). However, as our analysis showed, human-centered design is a framework-wide concern impacting many principles in each framework; it does not reside in an individual principle like "beneficence" or "non-maleficence"

(as it is portrayed in Jobin et al. 2019 and Floridi and Cowls 2019, respectively).

We observed that these frameworks use human-centered design focus in order to address concerns about human autonomy, equality, and uniqueness. For some scholars (Davis 2010; Steiner 2010; Bennett 2010), these concerns might appear protective and defensive, cultivating an egotistical perception of humanity. However, we view these concerns as legitimate within AI's formative period, because a human-centered design focus can ward off the worst visions of future AI and can set broad positive terms of AI success.

That said, from our perspective, a human-centered design focus cannot meaningfully be sustained without a corresponding environmental conscientiousness. The main risk of a human-centered design focus, after all, is that it does not necessarily take an environmental perspective (Steiner 2010). Many times, in fact, being human-centered has involved rejecting the claims that animals, plants, and wider environmental ecosystems are worth consideration (Steiner 2010). Thus, we are particularly drawn to frameworks' support of environmental well-being in their human-centered AI design focus, such as in BAAI and IEEE.

With this in mind, we can deepen how AI frameworks incorporate and implement an environmentally sensitive view of "human-centered" AI design. IEEE begins this work by noting that Shinto philosophy does not have a hard divide between humans and machines (p. 57). IEEE also notes that there are many legal alternatives for theorizing human-AI interaction short of personhood, such as "the treatment of pets, livestock, wild animals, children, prisoners, and the legal principles of agency, guardianship, and powers of attorney" (p. 255). Coeckelbergh (2010) extends some of these legal points of comparison to our experience of AI as well. One of the benefits of an environmental view of human-centered design is that specialists in animal ethics can provide helpful insights to building ethical models of human–AI interaction. These insights range from the idea that we have "indirect duties" to animals/AI (treating them well not for their sake but for our own), to the idea of assessing our responsibilities to animals/AI based on their varying capacities (Steiner 2010).

4.2 Alternatives to abstraction-oriented ethical reasoning

The second feature identified through our analysis is that most of the frameworks make extensive use of abstraction-oriented ethical reasoning. This type of reasoning may seem natural, helpful, and even unavoidable. For instance, Jobin et al. reflect an abstraction-oriented approach in their assumption that any recommendations for AI will be "derived" (p. 391) from principles developed in a framework. More broadly, using agreed-on abstract principles like "fairness" or "trustworthy" AI can provide points of common appeal, or *topoi*, that unite diverse stakeholders and help to guide deliberations for smaller details (McGee 1980).

However, scholarly work to revisit this assumption for AI has already begun, suggesting that abstraction-oriented reasoning is fragile. For Hagendorff (2020), corporate and individual practices expose that abstract ethical principles are typically disregarded by tech developers because many AI frameworks lack formal accountability mechanisms (and thereby are likely ineffective). Moreover, drawing on a feminist ethics of care, he argues that abstraction-oriented ethical reasoning is a masculine endeavor that ignores interpersonal dynamics like care, empathy, and nurture. Thus, in a compelling move, Hagendorff (2020) points to virtue ethics, which, he suggests, will help people to internalize a more ethically conscientious self-understanding and actually create an ethical path that liberates rather than constricts.

We supplement Hagendorff's (2020) calls for virtue ethics in two ways. First, we observe that BAAI is an exception to abstraction-oriented reasoning that also is not legible within the virtue ethics tradition, signaling the value of looking at different ethical and philosophical traditions for insight. Second, we identify a trend in rhetorical scholarship to highlight narratives (Young 1996) as an alternative to abstraction-first ethical reasoning. Narratives contextualize principles (reducing sometimes-strategic miscommunication) and embed insights at the level of practices. Although many of the frameworks in our analysis do make use of stories (especially Asimov), they do not allow the narratives to influence the frameworks' abstract ideals, revealing that merely including narratives in AI frameworks is not enough to access these benefits. Overall, we advocate a reflective equilibrium process that incorporates people's narratives/experiences and creates less dependence on abstract principles to guide AI development.

According to scholars of critical race theory, narratives incorporate people's rooted, specific knowledge (Crenshaw et al. 1996). Personal narratives have the power to challenge the "supposedly objective point of view [that] often mischaracterizes, minimizes, dismisses, or derides without fully understanding opposing viewpoints" (Delgado 1989). Moreover, stories can often reveal "that what we believe is ridiculous, self-serving, or cruel" in ways that objective discussion cannot (Delgado 1989). In addition, narratives can reveal fault lines within an abstract concept and societal structures themselves. To a great extent, Asimov's fictional narratives do this work by presenting different circumstances in which the Laws themselves are revealed as more complex in practice than the abstract framing of them would imply.

Narratives can also serve as sources for ethical reasoning. For Western Apache peoples, narratives rooted in certain places can even function as an ethical logic (Basso 1996). Places hold wisdom like water and people can learn to

"drink" wisdom from them through attention to their stories, becoming self-possessed, magnanimous, and watchful (p. 139–140). Such place-based narratives are grounded within physical spaces, granting them a much more concrete relevance. Thus, when applied to AI, these types of grounded narratives challenge the use of decontextualized, disembodied metaphors for conceptualizing and discussing such broad concepts like AI and "the cloud." For instance, Basso might suggest that we interpret Amazon's EC2 data centers as "places" for ongoing ethical reflection instead of conceptualizing "the cloud" as immaterial and abstract data.

A related perspective may be shifting towards more casuistic reasoning for future frameworks. In the law, casuistic reasoning begins with cases with clear-cut decisions, in order to identify and argue for similarities and differences that would give insight to a more difficult case (Jonsen and Toulmin 1988). Rather than applying generalized principles to every conflict or dispute, then, a casuistic approach looks at the specific circumstances of each case to gain a better understanding of the issue at hand. Another alternative to abstraction-oriented ethical reasoning is a "bottom-up" study of existing practices (Baba 1989; Schwartzman 1993; Woolgar and Latour 1986). This kind of analysis can reveal the enacted, yet oftentimes overlooked values and principles that already guide people's daily habits and motives, while still operating in a "scientific" mode, as some frameworks demand (IEEE, p. 172).

4.3 Redistributing authorship

The third feature identified through our analysis is that each of the frameworks has privileged authorship in terms of global and professional status. Hagendorff (2020) identifies that the authors of AI frameworks are generally male, and extensively analyses this in relation to the overt technicality of the frameworks. Other scholarly critiques of AI frameworks also identify a lack of authors from globally underrepresented regions like Asia and Africa (Floridi and Cowls 2019; Jobin et al. 2019) and attribute this lack to global power dynamics and regional economic development. Overall, these critiques of who is authoring AI frameworks echo Gillespie's point that we "must not conceive of algorithms as abstract, technical achievements, but must unpack the warm human and institutional choices that lie behind these cold mechanisms" (Gillespie 2014, p. 169).

Rhetorical scholarship about authorship allows us to include race, as well as soberly recognize the challenging work required for privileged people to shift their habits. Chakravartty et al. (2018) conduct an authorship study of the authors, editorial boards, and topics of published articles in the field of Communication Studies. They find that racism is perpetuated even through the publication rates, citation rates, and editorial positions of non-White Communication scholars (p. 254), in that underrepresentation in these areas is not race-neutral but oppressively distributes scholarly resources along racial lines. They find that a journal's relatively high percentage of non-White first authors is statistically associated with that journal's publication of more articles related to race, "suggesting that increasing non-White scholars' representation in the field is necessary to increase theorization and empirical research on race-related phenomena" (p. 259). Moreover, despite the commonly held assumption that citation practices should only be based on topical relevance and quality, they argue that citations "produce a hierarchy of visibility and value" (p. 257), in which non-White authors are continually left out despite producing high-quality and topically relevant work. Thus, Chakravartty's et al. (2018) reveals tangible consequences of authorship that speak to an oppressive past and existing systemic discrimination, whether or not texts themselves contain overtly racist content. These findings extend beyond race. For our paper, they suggest that privileged authorship of AI ethics frameworks can perpetuate the same "structures of powers" that support systemic oppression.

Once we acknowledge that authorship matters, we still must face the challenge of changing our collective habits. Nevertheless, we cannot overcome this challenge by merely satisfying a diversity quota, checking a checkbox, or even bringing underrepresented voices "to the table," so to speak (Ahmed 2012). In fact, from this light, the gestures toward inclusive authorship and various forms of feedback by several frameworks described in part 3.3 could even be uncharitably interpreted as doing just enough to avoid more significant change. In Google's case in particular, self-designed frameworks help structure or even avoid formal legislation into AI development and use (Mager 2012, p. 15).

Efforts to redistribute authorship take a significant amount of work. After all, many people in power have not developed the conceptual tools to deal with how deeply oppression has structured our world (Corrigan 2019). And scholars must "renounce their privilege" as part of responding to today's oppression (Wanzer 2012 pp. 654), yet this is complex and world-shifting. However, if we as a society are willing to put in the technical work necessary to make AI systems exceptional (and these frameworks suggest that we are), then we should also be willing to put in the redistributional work required to make AI systems truly beneficial for all.

4.4 Toward historically rooted justifications

The fourth and final feature identified through our analysis is that all the frameworks justify the need for their guidelines ahistorically. Other scholarship is silent on this front, prioritizing recently published frameworks and having little analysis, if any, of the historical factors at play in each

framework's depiction of risks from AI (Hagendorff 2020; Floridi and Cowls 2019; Jobin et al. 2019). Thus, an ahistorical view may even seem like common sense, like creating a clean slate to move forward. However, several scholars have examined how ahistorical justifications negatively impact social developments. Therefore, we suggest that the AI frameworks ignore opportunities for self-reflection and criticisms related to how we have gotten where we are today technologically. Because of this, structural changes appear easier than we might more realistically believe, and we are likely to be correspondingly naive or easily wearied by the task at hand.

In Safiya Noble's (2018) work, justifying reform by referencing history is important for understanding big tech's persistent racial *faux pas*. In a set of case studies on Google, from searches that return porn for "black girl" and return neo-Nazi propaganda for "black on white crimes," Noble identifies that Google's search algorithm has treated people of color worse than White people. Without historical attention, we are likely to see these as no more than a series of unfortunate accidents. Crucially, Noble theorizes these with reference to a particularly damaging historical practice of discrimination in the US: redlining. Formalized in the 1930s as a set of marked out physical spaces within a city, redlining persisted in the US for decades and contributed to massive Black–White wealth inequality. For Noble, physical redlining may have mostly ended, but Google's infrastructure has brought redlining into a new era: "digital redlining" (p. 10), or "technological redlining" (p. 1). Big tech's practices do not just instantiate oppressive social/economic interactions, they "reinforce" them (p. 10) and, therefore, demand our self-reflection, intentional rehabituation, and sustained attention to structural impacts of AI such as wealth inequality.

Ruha Benjamin (2019) shows another example of the importance of historical technological analysis. Each chapter of her book draws on snapshots of AI technologies in order to link AI failures to historical patterns. For instance, she develops an analogy between technological products ("code") with another US system: Jim Crow laws. These were a network of laws from the early and mid-twentieth century that effectively denied Black Americans of rights that they supposedly had according to the US Constitution. Jim Crow laws have already been extended to understand how the US prison system creates a "New" Jim Crow (Alexander 2010). Benjamin riffs on this by calling today's technological developments the "New Jim Code." Like Noble, Benjamin highlights that these do not appear *sui generis*, but "reflect and reproduce existing inequalities" (p. 5). It is naive to think AI missteps are isolated.

Applied to frameworks for regulating AI, frameworks that are exclusively forward-looking run the risk of promoting an erasure of the history of discrimination. They miss the opportunity to undo the damage of problematic historical trends by learning from them. As James Baldwin, a poignant commentator on the US Civil Rights Movement, puts it: "The great force of history comes from the fact that we carry it within us, are unconsciously controlled by it in many ways, and history is literally *present* in all that we do" (qtd. in Benjamin 2019, p. 10). Whether we recognize it or not, then, history surrounds us. Additionally, thinking historically is not opposed to technical specifications. As Rieder (2012) shows in tracing precursors to Google's PageRank algorithm, historical analysis sometimes requires a strong technical understanding.

Interestingly, some less high-profile AI frameworks resist the mainstream on this point. The Toronto Declaration, for instance, not only admits that "correcting for discrimination" is a necessary, ongoing process (Access Now 2018, Sect. 47a) that requires "proactive" measures (Access Now 2018, Sect. 36), but also advocates that companies disclose "the risks and specific instances of discrimination the company has identified" (Access Now 2018, Sect. 51a). These helpfully signal that risks from AI sometimes come from developers and corporations themselves.

5 Conclusion

This paper has worked to identify and delimit four of the features of an emerging AI mainstream from five high-profile AI frameworks. Our analysis has revealed a complex set of ambiguities that are worth evaluating. First, we supported a human-centered design focus that includes environmental conscientiousness and we show possibilities for applying insights from the animal-ethics tradition. Then, we pointed to abstraction-oriented ethical reasoning's lack of formal accountability mechanisms and suggest the following as alternatives: virtue ethics, holistic ethics (à la BAAI), narratives, casuistic thinking, and bottom-up studies. We also emphasized the need for awareness of authorship's consequences to meaningfully redistribute oppressive authorial and citational practices. Finally, we gestured towards the often-ignored relationship between historical trends and technological development to call for a more historically grounded exigence for AI frameworks.

Our analysis has indicated that the scope of the AI mainstream is wider than we may have thought, as seen by how easily Asimov's Laws have matched the features of today's AI mainstream. At a broader level, we note that many of the frameworks we examined are more dependent on popular conceptions than they would admit. As one example, the three rhetorical functions that we identified of human-centered design (i.e., maintaining human autonomy, human equality, and human uniqueness) can also be understood as responses to three collective fears ubiquitous in science fiction: against AI taking control, powerful

people using AI to wreak havoc, and AI infiltrating human society. That is, imaginative media are working out—and might even be said to be collaboratively constructing—the fears that serious AI developers are also engaging with. As a result, "high-profile" pop culture depictions of AI may have an important part to play in the emerging AI mainstream, both for the public and experts.

Overall, the concept of an AI mainstream has encouraged us to develop practical wisdom on these points. Future research can use the conceptual benefits of interpreting these features as part of a mainstream. These benefits include: handling ambiguity with ease, signaling the existence of alternatives, alluding to the power dynamics of circulation, and producing a healthy skepticism toward trends in AI regulation.

Acknowledgements Thanks to Lacey Davidson, Kristopher M. Torres, and Mary Glavan for providing us with helpful feedback. Thanks to the Living with AI Fall 2018 seminar for initial feedback and ideas for the direction of this paper. Thanks to Michael Hemenway for pointing us to recently published comparisons of AI frameworks.

References

Access Now (2018) The Toronto declaration: protecting the rights to equality and non-discrimination in machine learning systems.
Ahmed S (2012) On being included: racism and diversity in institutional life. Duke University Press, Durham
AI HLEG (2019) Ethics Guidelines for Trustworthy AI. Retrieved from High-Level Expert Group on Artificial Intelligence
Anderson MR (2017) After 75 Years, Isaac Asimov's Three Laws of Robotics Need Updating. The Conversation. https://theconversation.com/after-75-years-isaac-asimovs-three-laws-of-robotics-need-updating-74501
Asimov I (2004 [1950]). *I, Robot* (Vol. 1). Spectra
BAAI (2019) The Beijing AI Principles. Beijing Academy of Artificial Intelligence
Baba M (1989) Organizational culture: revisiting the small-society metaphor. Anthropol Work Rev 10(3):7–10
Balkin J (2017) The three laws of robotics in the age of big data. Ohio State Law J 78:27
Basso KH (1996) Wisdom sits in places: landscape and language among the western apache. UNM Press, Albuquerque
Benjamin R (2019) Race after technology: abolitionist tools for the New Jim Code. John Wiley & Sons, Hoboken
Bennett J (2010) Vibrant matter: a political ecology of things. Duke University Press, Durham
Burke K (1966) Language as symbolic action: essays on life, literature, and method. Univ of California Press, Berkeley
Chakravartty P, Kuo R, Grubbs V, McIlwain C (2018) #CommunicationSoWhite. J Commun 68(2):254–266
Chia RG (2019) Despite reportedly tracking and ranking its own citizens, China Now Says AI Research Should Respect People's privacy. Business insider. https://www.businessinsider.my/despite-reportedly-tracking-and-ranking-its-own-citizens-china-now-says-ai-research-should-respect-peoples-privacy
Coeckelbergh M (2010) Humans, animals, and robots: a phenomenological approach to human-robot relations. Int J Social Robot 3(2):197–204. https://doi.org/10.1007/s12369-010-0075-6
Corrigan L (2019) Decolonizing Philosophy and Rhetoric: dispatches from the Undercommons. Philosophy Rhetoric 52(2):163. https://doi.org/10.5325/philrhet.52.2.0163
Crenshaw K, Gotanda N, Peller G, Thomas K (eds) (1996) Critical race theory: the key writings that formed the movement. The New Press
Davis D (2010) Inessential solidarity: rhetoric and foreigner relations. University of Pittsburgh Press, Pittsburgh
Dean J (2019) Google AI blog: looking back at Google's research efforts in 2018. Google Blog. https://ai.googleblog.com/2019/01/looking-back-at-googles-research.html
Delgado R (1989) Storytelling for oppositionists and others: a plea for narrative. Mich Law Rev 87(8):2411. https://doi.org/10.2307/1289308
Dvorsky G (2014) Why Asimov's Three Laws of Robotics Can't Protect Us. Gizmodo. https://io9.gizmodo.com/why-asimovs-three-laws-of-robotics-cant-protect-us-1553665410
Floridi L, Cowls J (2019) A unified framework of five principles for AI in society. Harvard Data Science Review
Flower L (2008) Community literacy and the rhetoric of public engagement. SIU Press, Carbondale
George SK (2017) Total enframing: global south and techno-developmental orthodoxy. AI Soc 32:191–199
Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski P, Foot K (eds.) Media Technologies: Essays on Communication, Materiality, and Society, 167
Hagendorff T (2020) The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 30(1):99-120
IEEE (2016) Ethically aligned design. IEEE Standards v1
Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. Nat Machine Intell 1(9):389–399
Johnson AE (2012) The forgotten prophet: Bishop Henry McNeal Turner and the African American prophetic tradition. Lexington Books
Jonsen AR, Toulmin S (1988) The abuse of casuistry: a history of moral reasoning. University of California Press, Berkeley
Knight W (2019) Why does Beijing suddenly care about AI ethics? MIT Technology Review. https://www.technologyreview.com/2019/05/31/135129/why-does-china-suddenly-care-about-ai-ethics-and-privacy/
Mager A (2012) Algorithmic ideology: how capitalist society shapes search engines. Inform, Commun Soc 15(5):769–787
McGee MC (1980) The "Ideograph": a link between rhetoric and ideology. Q J Speech 66(1):1–16
Noble SU (2018) Algorithms of oppression: how search engines reinforce racism. New York University Press, New York
Penman W (2018) A field-based rhetorical critique of ethical accountability. Q J Speech 104(3):307–328
Pichai S (2018) AI at Google: our principles. Google blog. https://www.blog.google/technology/ai/ai-principles/
Proyas A (2004) I, Robot [action, crime, drama, sci-fi, thriller]. Twentieth Century Fox, Mediastream Vierte Film GmbH & Co. Vermarktungs KG, Davis Entertainment. IMDb
Ratcliffe K (2005) Rhetorical listening: identification, gender, whiteness. SIU Press
Responsible AI Practices (2018) from Google AI website: https://ai.google/responsibilities/responsible-ai-practices/. Accessed 11 July 2019
Rieder B (2012) What is in PageRank? A historical and conceptual investigation of a recursive status index. Computational Culture (2)
Roberts-Miller P (2007) Deliberate conflict: argument, political theory, and composition classes. Southern Illinois University Press, Carbondale
Schiappa E (2012) Defining marriage in California: an analysis of public and technical argument. Argum Advocacy 48(4):216–230

Schwartzman HB (1993) Ethnography in organizations. Sage 14(4):614
Singer PW (2009) Wired for war: the robotics revolution and conflict in the 21st Century. Penguin
Steiner G (2010) Anthropocentrism and Its discontents: the moral status of animals in the history of western philosophy. University of Pittsburgh Press, Pittsburgh
Walker K (2018) Google AI Principles Updates, Six Months in. Google Blog. https://www.blog.google/technology/ai/google-ai-principles-updates-six-months/
Wanzer DA (2012) Delinking rhetoric, or revisiting McGee's fragmentation thesis through decoloniality. Rhetoric and Public Affairs 15(4):647–657
Woolgar S, Latour B (1986) Laboratory life: the construction of scientific facts. Princeton University Press
Young IM (1996) Communication and the other: beyond deliberative democracy. In: Benhabib S (ed.), Democracy and difference: contesting the boundaries of the political, 120–135