JOURNAL OF SOCIAL COMPUTING
ISSN 2688-5255 05/06 pp266−283
Volume 2, Number 3, September 2021
DOI: 10.23919/JSC.2021.0031

From Ethics Washing to Ethics Bashing: A Moral Philosophy View on Tech Ethics

Elettra Bietti*

Abstract: Weaponized in support of deregulation and self-regulation, "ethics" is increasingly identified with technology companies' self-regulatory efforts and with shallow appearances of ethical behavior. So-called "ethics washing" by tech companies is on the rise, prompting criticism and scrutiny from scholars and the tech community. The author defines "ethics bashing" as the parallel tendency to trivialize ethics and moral philosophy. Underlying these two attitudes are a few misunderstandings: (1) philosophy is understood in opposition and as alternative to law, political representation, and social organizing; (2) philosophy and "ethics" are perceived as formalistic, vulnerable to instrumentalization, and ontologically flawed; and (3) moral reasoning is portrayed as mere "ivory tower" intellectualization of complex problems that need to be dealt with through other methodologies. This article argues that the rhetoric of ethics and morality should not be reductively instrumentalized, either by the industry in the form of "ethics washing", or by scholars and policy-makers in the form of "ethics bashing". Grappling with the role of philosophy and ethics requires moving beyond simplification and seeing ethics as a mode of inquiry that facilitates the evaluation of competing tech policy strategies. We must resist reducing moral philosophy's role and instead must celebrate its special worth as a mode of knowledge-seeking and inquiry. Far from mandating self-regulation, moral philosophy facilitates the scrutiny of various modes of regulation, situating them in legal, political, and economic contexts. Moral philosophy indeed can explainin the relationship between technology and other worthy goals and can situate technology within the human, the social, and the political.

Key words: ethics; technology; artificial intelligence; big tech; ethics washing; law; regulation; moral philosophy; political philosophy

1 Introduction

On May 26th, 2019, Google announced that it would put in place an external advisory council for the responsible development of AI, the Advanced Technology External Advisory Council (ATEAC).[1] Following a petition signed by 2556 Google workers demanding the removal of one of the body's board members, anti-LGBT advocate Kay Coles James, the advisory body was withdrawn approximately one week after its announcement.[2, 3] On December 3rd, 2020, Timnit Gebru, a Google AI researcher, was abruptly fired for sending an internal letter to Google employees which discussed her superiors' questionable resistance to the publication of a research paper she co-authored.[4–6] Her Tweet produced a wave of reactions in academia and beyond, with many Google employees subsequently quitting.[7] These episodes and the backlash they produced provide a salient illustration of the tensions around the corporate use of "ethics" language in technology circles. Corporate and policy instrumentalization and misuse of such language in technology policy have taken two forms.

• Elettra Bietti is with the Harvard Law School, Harvard University, Cambridge, MA 02138, USA. E-mail: ebietti@sjd.law.harvard.edu.
* To whom correspondence should be addressed.

Manuscript received: 2021-05-20; revised: 2021-11-19; accepted: 2021-11-25

On one hand, the term has been used by companies as an acceptable façade that justifies deregulation, self-regulation or market driven governance, and is increasingly identified with technology companies' self-interested adoption of appearances of ethical behavior. Such growing instrumentalization of ethical language by tech companies has been called "ethics washing".[8] Beyond AI ethics councils or AI Ethics researchers, the ethics washing critique extends to corporate practices that have tended to co-opt the value of ethical work: the hiring of in-house moral philosophers who have little power to shape internal company policies; the careful selection of employees that will not question the status quo; the focus on humane design—e.g., nudging users to reduce time spent on apps—that does not address the risks inherent in tech products themselves;[9] the funding of "fair" machine learning systems combined to the defunding of work on algorithmic systems that questions the broader impacts of those systems on society.[10, 11]

On the other hand, the technology community's criticism and scrutiny of instances of ethics washing, when imprecise, have sometimes bordered into the opposite fallacy, which the author calls "ethics bashing". This is a tendency, common amongst non-philosophers, to simplify the issues around tech "ethics" and "moral philosophy" either by drawing a sharp distinction between ethics and law and defining ethics as that which operates in the absence of law[12] or by conflating all forms of moral inquiry with routine politics, for instance by merging or drawing artificial separations between the frameworks of "ethics", "justice", and "political action".[13, 14] Distinguishing between "law" and "ethics" is a common legal positivist move, configurable within a long philosophical tradition that sees the practice of making, interpreting, and applying law as processes whose existence and relevance are distinct and separable from their moral and societal implications.[15] The relation between "ethics", "justice", and "political action" instead is complex. Understanding ethics and moral inquiry as either a mode of political action or a discrete, individual-centric, and particularized exercise that is easily instrumentalized and is unsuited to tackling political and institutional questions is misleading yet frequent. As described by Jacob Metcalf, Emanuel Moss, and Danah Boyd, the distinction between narrow "ethics" and capacious "justice" became a central focus of discussions during the 2019 ACM Conference on Fairness, Accountability and Transparency.[13]

Equating serious engagement in moral argument with the social and political dynamics within ethics boards or understanding ethics as a methodological stance that is antithetic to—instead of complementary to and inherent in—serious engagement in law-making and democratic decision-making, is a frequent and dangerous fallacy. The misunderstandings underlying the broad trend of ethics bashing are at least three-fold: (1) philosophy is either confused with "self-interested politics" or understood in opposition to law, justice, political representation, and social organizing; (2) philosophy and "ethics" are seen as a formalistic methodology, vulnerable to instrumentalization and abuse, and thus ontologically flawed; and (3) engagement in moral philosophy is downplayed and portrayed as mere "ivory tower" intellectualization of complex problems that need to be dealt with through alternative and more practical methodologies.

Grappling with the role of ethics in tech policy requires moving beyond both ethics washing and ethics bashing and seeing ethics as a mode of inquiry that informs work in law, policy, and technological design alike in emancipatory directions. Policy-makers, lawyers, technologists, corporates, and academics do moral theorizing all the time. Asking whether a corporate ethics council can improve internal policy-making, whether a given machine learning system can lead to fairer criminal justice enforcement, or whether a given corporate decision to fire a researcher or ban facial recognition is acceptable in context involves asking moral questions that, if properly framed, can lead to a better understanding of these phenomena and also to better policies. Awareness of the ubiquity of morality would enable all actors in the technological and AI space to contextualize their work with greater subtlety, at several levels of abstraction, and to more rigorously assess the legitimacy of corporate self-regulation and other ethics initiatives.

One aim of this article is to distinguish between what ethics is often thought to be (a neutral and context-independent methodology, a self-interested corporate rhetoric) and what ethics could be (a principled methodology for evaluating political disagreements around technology). To understand that distinction, another distinction must be captured between the intrinsic and the instrumental value of ethics. The intrinsic perspective sees ethics as a mode of inquiry which is independently valuable as an aspirational process, particularly for those engaging in it. The instrumental perspective instead sees the value of ethics as lying in its results. The value of ethics understood in this way depends on its end-results, ethics' causal role in bringing about desired results, such as reputation, innovation, and profit. Intrinsic and instrumental perspectives on ethics and moral inquiry are not mutually exclusive. One can understand ethics as an intrinsically valuable process with valuable results. However, distinguishing facial appearances of ethics from approaches that emphasize ethics' potential entails emphasizing intrinsic value over instrumental value. The author will argue that the more the process of engaging in ethics is motivated by outcomes independent of the process itself—the less ethics is taken as an intrinsically valuable process—the weaker its moral value becomes for society. Ethics washing and ethics bashing are instrumental understandings of ethics, in that both positions or tendencies envision or experience ethics as a means to an end and nothing more.

What is at stake in recent controversies around the weaponization of "ethics" rhetoric are also competing moral conceptions of technology companies' role. Corporate-friendly conceptions benefit from inserting ethical work within larger communications and public relations strategies.[13, 16−18] Critical conceptions reject these corporate efforts and prefer participatory democracy and activism.[11, 19] Yet both corporations and their critics obscure the potential role that moral inquiry can and must play in developing a thicker conception of technology politics. There is no neutral perspective "outside morality" from which the normative implications of technology can be teased out. It should thus be possible to maintain a critical outlook on the instrumentalization of ethics in technology settings, while also recognizing the special value and centrality of moral inquiry to expanding horizons.

This article has two goals. First, it aims to articulate the weaknesses of both the ethics washing and ethics bashing fallacies, explaining why both are impoverished views of the relationship between technology and ethics. Second, it aims to clarify the potential of moral philosophy in debates about the impact of new technologies on society and thereby to dissipate misunderstandings of moral philosophy as either too abstract to inform concrete policy or as a red herring that prevents proper focus on political and social action. Far from constituting a barrier to appropriate governance, moral philosophy enables us to seriously scrutinize the future of technology governance, law, and policy, and to understand what humans need from new technologies and innovation from a unique vantage point.

The article is structured as follows. In Sections 2−4, the article begins by explaining the function and meaning of ethics and moral philosophy, some common criticisms of moral philosophy, and what it is for. Section 5 of the article then provides background on the rise of ethics in tech and the advent of so-called "ethics washing". In Section 6 it explains the limits of existing critiques of ethics washing, identifying "ethics bashing" as a fallacious depiction of ethics as opposed to law, politics, or justice. In Sections 7 and 8, adopting a view internal to moral philosophy, the author engages in a moral argument and shows that commitment to moral principles and engagement in moral reasoning also leads to the conclusion that corporate ethics efforts are by and large wrong and that ethics is antithetic to what happens inside corporate settings. Finally, Section 9 of this article suggests a way forward that moves beyond both ethics washing and ethics bashing, that adopts a less instrumentalist position on ethics, and that requires developing governance frameworks that enable the emergence of renewed moral, political, and legal thinking and action outside corporate settings.

2 Ethics and Moral Philosophy

The English word "ethics" is derived from the ancient Greek words ēthikós and êthos which refer to character and moral nature.[20] Morality comes from the Latin moralis which means manner, character, and proper behavior. Both "ethics" and "morality" thus refer to the study of good and bad character, appropriate behavior, and virtue. The two terms are often employed interchangeably but have slightly distinct uses and connotations. Morality is often associated with etiquette and rules of appropriate social behavior, whereas ethics has instead a more personal connotation. Ethics pertains to the cultivation of individual virtue abstracted from society and is sometimes used to refer to personal and professional standards of behavior embodied in "codes of ethics". In Confucian philosophy, morality is about respecting the family and pursuing social harmony and stability through virtues including altruism, loyalty, and piety.[21]

In the discussion to follow, the term "ethics" will refer to the rhetoric of morality employed in technology circles, and "moral philosophy" will instead refer to the philosophical discipline that investigates questions around human agency, freedom, responsibility, blame, and the relationships between individuals, amongst other questions. The author adopts a primarily Anglo-American liberal approach to the practice and understanding of moral philosophy[22] but the author's perspective is by no means intended to close the door to alternative approaches to moral philosophy and ethics. According to some accounts, moral philosophy's scope is limited to relationships between humans and ethics extends instead beyond humans to animals and nature. Some would also distinguish moral from political philosophy while others such as Ronald Dworkin see them as interconnected.[23] Like Ronald Dworkin, the author construes the "moral" widely as consisting of the domain of "value", i.e., an evaluative mode of inquiry which is distinguishable from scientific or descriptive modes of inquiry, which focus on facts.[23, 24] The domain of "value" is the specific domain of inquiry of moral philosophers.

To better illustrate what moral philosophy is, consider the example of surveillance. Let us ask: what is wrong or unethical about big data and certain forms of surveillance? Disparate arguments can be offered to show that big data and surveillance are wrong in some respects or worth carrying out in other respects. Different persons will likely have different views on which of these arguments are strongest. As philosophers might put it: the morality of surveillance is an evaluative matter, i.e., a matter on which reasonable people disagree because they hold competing moral interpretations of what is at stake. Numerous lines of reasoning support the wrongness of surveillance and business models that rely on data extraction. Surveillance is objectionable on self-development and virtue ethics grounds because it incentivizes self-censorship, reducing human beings' ability to develop themselves or to engage in other valuable causes for fear that these actions will be held against them. Another argument focuses on harm: some surveillance and big data activities cause harm to individuals (e.g., they lead to unjustified and stereotype-enhancing discriminatory treatment, they create asymmetries of knowledge and power, they perpetuate pre-existing and unjustified inequalities). A third line of reasoning focuses on equal dignity and respect for persons: some forms of data processing and surveillance fail to treat individuals as equally worthy of respect because they are covert and because some people are surveilled more than others.

Each line of argument entails a different way of evaluating policy. For instance, if someone considers that surveillance inhibits the pursuit of worthy behavior or individuality, they might be satisfied with aspects of big data and surveillance practices that enhance the pursuit of certain worthy life goals, including certain targeted and personalized work opportunities, as long as they are empowering and equally distributed. On the other hand, if one believes that the core problem is that the data collected can cause unintended harm to individuals, they might advocate for solutions that minimize discriminatory impacts and ensure that harms are reduced. Finally, someone who believes that surveillance and the opacity of big data activities are denials of respect for the persons surveilled might be keen to ban surveillance completely or to reduce any tolerable surveillance to a de minimis threshold.

Which reasons we find most weighty is a matter of commitment and deliberation on how to actualize moral values such as autonomy, equality, and human flourishing. The process of weighing some reasons against others allows us to overcome the intuitive belief that "surveillance feels creepy",[25] and to instead ground or re-evaluate one's commitment to privacy or its limitation based on carefully weighed argument on how different forms of surveillance and data extraction might interact with autonomy, dignity, equality, and human flourishing. Identifying the drawbacks of surveillance business models and their morally unacceptable core also facilitates the design of nuanced concrete strategies for addressing them.

This process of revising and refining moral beliefs through philosophical inquiry is what John Rawls has called reflective equilibrium.[26] What Rawls' methodology and other analogous modes of moral evaluation have in common is that they provide a lens through which to interpret issues of societal importance, to locate them within existing debates, consider them from all relevant standpoints, and evaluate which angle or way of approaching them is capable of shedding the most valuable light on the issues themselves. When engaging in this process, the broader the spectrum of considerations that are taken into account in moral theorizing, the more interesting, capacious, and morally significant are the outcomes, and the more inspiring and valuable are its practical implications.

It is also important to emphasize that moral philosophy and ethics can mean different things as part of different fields of study and intellectual traditions. The above is intended to capture only a glimpse of a larger roadmap of possible uses of the terminology of ethics and moral philosophy in technology governance and policy. It is not intended to fix the meaning of these rich and complex modes of inquiry.

3 What Moral Philosophy Is For

A key question is what ethics and moral philosophy are for and what they can contribute to existing technology policy debates. In asking this question, The author focuses on the reflexive value of engaging in moral reasoning from the perspective of those engaging in it, i.e., “from within”. In the technology policy context, moral and other philosophical work is valuable in at least four ways for those who pursue it.

First, philosophical reasoning and deliberation can provide a meta-level perspective from which to consider any disagreement relating to the governance of technology. Instead of taking arguments narrowly, intuitively, or personally, philosophical reasoning provides a framework for stepping back, situating any problem within its broader context and understanding it within or in relation to other relevant or analogous debates. As such, the practice or method of engaging in moral argument allows us to broaden our perspective and to look at a debate from a wider lens, overcoming confusions, filling in gaps, correcting inconsistencies, and drawing clarifying distinctions. In debates on the acceptability or necessity of facial recognition technologies, for instance, a philosophical method can help us rethink our reasons for rejecting or promoting existing technologies, clarify points of agreement between a variety of opponents to these technologies, and focus on where disagreements lie and what they entail in practice: what freedom, equality, and human flourishing require in an era of structural surveillance and systemic inequality. Otherwise put, philosophy is a good antidote to knee-jerk reactions: it can help reduce unbridgeable value conflicts and make agreement possible by moving discussions between different levels of specificity or abstraction. This is not to say that ideology and value conflicts are unimportant, but merely to recognize the importance of philosophy as a method aimed at overcoming or clarifying those conflicts.

A second, related, contribution of moral philosophy to tech debates is that it adds rigor principled thinking to value-laden, emotional, or subjective discussions. Moral philosophy should be understood as an explanatory mode of inquiry which requires us to set out the justifications and reasons for advancing one view and not a different one. By centering attention on the explanation and the justification for a position, philosophy enables a dialectic to take place, a Socratic dialogue which we can have internally with ourselves or externally with others, that sheds light on blind spots and enables fluid and iterative repositioning. Winning the argument is not as important as laying all its facets on the table. Such principled and disinterested inquiry is frequently absent in technology policy and governance discussions for at least two reasons. The first is that current policy debates are instinctive, emotional, polarizing and inimical to measured reflection. The second is that many of these debates are mediated by platforms whose corporate incentives are difficult to align with disinterested reflection on societal impacts.[27, 28]

Third, a normative philosophical lens can substantively move us beyond a narrow focus on procedural fairness, diversity, and representation in technology governance, and towards substantive goal evaluation. As explained in more detail below, the problem is not just whether an AI ethics board’s members have diverse perspectives and backgrounds, but also whether the board’s decisions can actually constrain Google’s profit-motivated actions. Similarly, the question is not just whether a facial recognition algorithm properly recognizes black faces, but whether such algorithm is deployed in circumstances where it can harm black people. A capacious moral philosophy approach can help us move beyond checklists and proceduralism to question whether an existing or future structural governance framework and its substantive outcomes are morally acceptable and worth pursuing.

Fourth, far from obscuring ideological conflicts and structural divisions[19, 29] engaging in moral philosophy can facilitate dialogue, encourage the building of common ground, and provide a basis for collaborative and participatory approaches to policy-making capable of bridging divides in a polarized landscape. An important drawback of critical work that centers on power, value conflicts, and unbridgeable ideological divides is that it renders dialogue between people holding different views or occupying different social positions more difficult. Pursuing such strategies has its advantages but it can also lead to fragmentation in an already polarized and emotions-driven public sphere. Understanding philosophy as a dialectic discipline that enables empathy and grounds methodology in the aspirational possibilities of commonality, justification, and conflict resolution can instead help navigate fragmentation and polarization today. The many "embedded ethics" initiatives at computer science and philosophy departments in the United States and beyond are fostering greater debates and have been shown to promote the building of common ground across disciplinary boundaries.[30−33]

Still, while acknowledging the important contributions of Western philosophy to the promotion of an inclusive and discursive public sphere, awareness of how power and inequality manifest within such discursive public sphere is key. Not every person has the same voice and the same ability to be heard.[34] Equalizing a space in the face of structural inequality must thus be one of the first considerations when building spaces for dialogue and "ethical" reflection. Contemporary approaches that embed ideology and structural power asymmetries within normative philosophical inquiry[19, 29, 35] account for the advantages of a discursive methodology while expanding the horizon of philosophical inquiry to include issues of structural inequality, power, domination, and ideological entrenchment.

4 How to Criticize Ethics and Moral Philosophy

Work in moral philosophy and ethics has a number of limitations. Before turning to the rise of ethics discourse in technology and the fallacies associated with that trend, here are six ways of criticizing moral philosophy that are targeted at moral philosophy as a reflexive exercise and as a methodology. By addressing these important criticisms, my aim is to shed light on moral philosophy as a critical method, showing that it can channel change, re-assessment, and revision of commonly held beliefs.

First, philosophy can be criticized for being abstract and for not being accessible to large audiences. This makes philosophical work often unsuited to advocacy or activism or to making provocative contributions to time-sensitive issues. Philosophy is also rarely suited to opeds, for example, or to those who aim at quick and easy policy fixes. Yet depth and abstraction are also one of the discipline's advantages: engaging in philosophical work prompts us to pause and think, to shield our thinking from pragmatic pressures, to enlarge the temporal and geographical scope of our research scope. As we engage in this process, our intuitions change, we extend our thoughts or revise them so that they can connect with and make sense of other problems, we learn how to think slower, to think with more depth and more systematically. To achieve meaningful cultural and social renewal in the technology industry, countering a technological culture of fast-paced permissionless innovation driven by an ethos of "move fast and break things", slowness needs to be taken more seriously.[36]

Second, some work in moral philosophy, particularly in its connections with technology, is seen as not going far enough prescriptively or as doing harm in practice. Recent work in social science, for example, has attempted to rely on the philosophical heuristic of the trolley problem[37] to address the regulation of Autonomous Vehicles (AVs), with scarce practical success and generating significant controversy. The Moral Machines experiment at MIT,[38, 39] a large-scale experiment that gamifies the trolley problem to extrapolate aggregate data and then guidelines for programming AVs, has been criticized for simplifying, scaling, and misusing a case-specific and contextual philosophical mode of reasoning.[40] Similarly, Basl and Behrends argued that attempts at applying trolley problem insights directly to AV policy are flawed because they fail to take into account the complexity and contextuality of machine learning development.[41]

More generally, entrenching high level principles for ethical AI in Codes[42] also arguably remains too abstract to guide individuals and policy-makers' actions in practice on AI questions.[11, 43, 44] In the absence of a deep understanding of context, focusing on the trolley problem or outlining high level theoretical principles for ethical AI appears unlikely to lead to workable and morally compelling regulatory strategies. These examples leave us perplexed: much philosophical work seems irrelevant or unsuited to resolving pressing problems in technological contexts. What is needed however is not less philosophical work, but more thinking on what moral principles can do in practice, and what they mean contextually. Helen Nissenbaum's work on contextual privacy is an important example of how thoroughly articulating the contextual implications of abstract privacy norms can impactfully guide the work of communities of practice.[45]

Third, the application of philosophical work can have effects in practice that sometimes contradict the philosopher's motivations. Hegel and Nietzsche's philosophical ideas have been instrumentalized by the German Nazi regime to pursue inhumane ends, an instrumentalization that had little connection to what these philosophers were actually doing or thinking.[46, 47] More concretely, philosophers frequently understand reflection and engagement with the politics and context of their work as corrupting, and thereby fail to prevent misuses of their ideas for unworthy ends. The hiring of moral philosophers by technology companies is but one instance in which philosophical ideas need to be scrutinized in context; such work cannot be taken at face value just because they are the ideas of a trained philosopher. Philosophers are hired, and then their skills are subordinated to the commercial goals of their employers. In this way, work that might have seemed apolitical in an academic setting acquires a new politics. This work can become harmful if it hides under the appearance of neutral thinking allowing the legitimation of controversial states of affairs, such as the secrecy of algorithms and their control by private companies. As important as it is, this criticism however should not be seen as fatal to the kind of work philosophers do. The emergence of in-house philosophers means philosophical work must be scrutinized with even greater care, must be publicly accountable, and philosophers must exercise an enhanced level of caution regarding the context and consequences of what they do. Importantly, the funding of philosophical work in the technology and governance field must be disclosed and discussed more openly.

Fourth, work in ethics can be understood as normalizing, as an attempt to discipline social life by devising and applying universally applicable norms of conduct that entrench existing power dynamics by placing them outside the realm of contestation.[48] Marxist critics of moral philosophy have also argued that capitalist incentives can influence philosophical work in directions that favor the interests of businesses and elites.[49] Ethnographers speak of "ordinary ethics" as the descriptive way ethics and morality structure routine social interaction.[13] Zigon however emphasizes the importance of distinguishing routine and unconscious moral claims from conscious ethical claims that arise during "breakdown" moments and are aimed at changing a culture and at "returning to the unreflective mode of everyday moral dispositions".[50] While Zigon's anthropological perspective on morality and ethics captures the pivotal role played by moments of breakdown and moral dilemma, he still sees morality and ethics as fundamentally about the need to return to unreflected normality, to revise beliefs so they can be fixed, routinized, and remain unchallenged once again. For philosophers, instead, morality and ethics are centrally about reflectiveness, conscious revising of beliefs and constant changes to the status quo. Contrary to anthropologists and ethnographers, moral philosophers and ethicists are only marginally concerned with the normalization of moral beliefs. For a philosopher, the task is indeed to engage in direct moral questioning about these beliefs and to bring them to the foreground of our consciousness, instead of emphasizing their regularities and embeddedness in social norms and cultural contexts.

Fifth, philosophical theorizing is frequently criticized for creating an appearance of principled reasoning, neutrality, and objectivity when much of what is at play are a philosopher's subjective views.[19, 51] There is some validity to this criticism, but it is less powerful than it first appears. Good normative philosophical work does not attempt to convey an appearance of absolute objectivity. Quite the contrary, such work is very clear regarding the uncertain bases on which it stands. A large share of Anglo-American moral philosophy follows Rawls' reflective equilibrium or a similar method, to progressively match intuitions and beliefs to considered judgments. This iterative process is one of many approaches that Anglo-American philosophers use to formulate normative conclusions. Although any philosophical conclusion necessarily originates in a thinker's subjective intuitions and beliefs, it is also the product of structured and iterative revisions. It gives conclusions a normative weight or subtlety that raw intuitions do not have. Far from presenting ultimate and final words on a subject, good philosophical work is rigorous yet porous and open to scrutiny: its aim is to broaden perspectives, allowing us to see the limits of the existing and to constantly revise our beliefs.

Finally, sociologists have argued, often rightly, that philosophy is not sufficiently from a gender and racial perspective in particular, dominated instead by Western male figures.[52]

These criticisms are grounded in the idea that moral philosophy can be a worthy enterprise but that its objective appearance or moral weight too often leads philosophers in the wrong direction. Philosophers and theorists interested in the potential of ethical reflection in technology should not only be aware of these vulnerabilities but must also combat them by embedding inclusion and resistance to the exploitation and instrumentalization of moral inquiry into their very methodologies and practices.

As shown, moral philosophy is a reflexive pursuit that is valuable as a process for those who engage in it in view of making sense of the world around them with caution and empathy. Moral philosophy in this sense is not a synonym of the ethical initiatives that occur within corporate settings which are mostly self-centered and instrumental;[18] it is an exercise that, if construed radically as an inclusive emancipatory methodology, is in inherent tension with industry players' profit logics. In Section 5, the author explains the development and rise of technology ethics and its entrenchment within private companies, a trend often aimed at reputational enhancement which has been called "ethics washing".[8]

5 The Rise of Tech Ethics and Ethics Washing

In an important essay in 1980, Winner showed that artifacts have politics in two important ways: technologies embed and express the biases and power relations of the society and people who design them, and the deployment and use of these artifactual affordances in turn change and shape the politics and power relations in society.[53] The rise and promise of machine learning and artificial intelligence technologies have brought about a renewed urgency to the debate on the political nature of technology and its ethical implications. A number of prominent books and articles on the subject have shown that the deployment of artificial intelligence can have significant consequences for privacy, human dignity, equality and non-discrimination, gender, social, racial, and economic justice.[54–61] The growing awareness of AI's societal implications and political nature, and a significant "techlash",[62] have led companies involved in developing AI systems to pay attention to the ethical implications of data science and artificial intelligence.

In the last few years technology ethics has grown in popularity and been adopted and endorsed in a multitude of overlapping forms.[43] High-level statements of principled artificial intelligence have been created or endorsed by private companies, civil society, governments, as well as transnational and multi-stakeholder entities.[42] Ethics training has been developed and embedded in the computer science curriculum of a growing number of universities.[30–32, 63] The growing research field of AI and the growing body of research around its ethical and societal implications has led to the creation of a number of new conferences and dedicated research institutes.[42]

Private companies have been involved in these efforts at each level: developing and publicly sharing statements of AI principles,[42] hiring in-house ethicists,[64] forming ethics councils and bodies,[3] and putting in place ethics and diversity trainings and structures for their employees.[18] As regards principles, Google, for instance, has published principles emphasizing the need for AI applications to be socially beneficial, to avoid creating or reinforcing bias, to be safe and accountable.[65] Microsoft and IBM have also engaged in codifying principles and procedures for safe and trustworthy AI.[66, 67] Microsoft's website states the need to move beyond principles and toward implementation of ethical AI through ad hoc internal bodies:

> We put our responsible AI principles into practice through the Office of Responsible AI (ORA) and the AI, Ethics, and Effects in Engineering and Research (Aether) Committee. The Aether Committee advises our leadership on the challenges and opportunities presented by AI innovations. ORA sets our rules and governance processes, working closely with teams across the company to enable the effort.[67]

When they do not engage directly in crafting statements of principles and setting up internal ethics boards, private companies sponsor AI conferences, research institutes and efforts that shape the research agenda and discourse around the societal impact of AI.[68] The Partnership on AI, a non-profit established to study and formulate best practices on AI technologies, was founded by Amazon, Facebook, Google, DeepMind, Microsoft, and IBM, and is entirely funded by industry stakeholders. Palantir, Google, and Facebook frequently fund major law, computer science, and privacy conferences.[18, 43] In turn, AI ethics is becoming a business, with consultancy firms and law firms developing AI ethics expertise to assist tech companies in their compliance efforts.[69, 70]

As these instances show, companies such as Google, Facebook, Microsoft, and Palantir are concerned about their ethical reputation in the face of new technological developments in data science and beyond. Their efforts to promote and arguably build more trustworthy and ethical AI indicate a calculative stance, a method for preempting financial and reputational risk, more than a recognition of the political nature of AI and its implications.[13, 14, 16] Even though it might be argued that the intentions behind these initiatives are good, the practices themselves are too limited and opportunistic to be in line with a conception of morality and ethics as reflexive capacious exercises that can foster disinterested selfless change. Overall, speaking of AI "ethics" instead of AI "politics" can be seen as a way to depoliticize and normalize the impacts of company efforts in this space,[14] allowing companies to "ethics wash" their reputations and to narrow the space for real debate and change in AI.[8, 71]

6 Critiques of Ethics Washing: Merits and Limits

Efforts such as embedding ethicists or ethical guidelines within industry practices and creating codes of ethical principles aimed at more responsible and trustworthy technological design have been criticized by scholars for normalizing and depoliticizing data science and AI (Green, this issue). They have been criticized for bringing about a performative "transformation of ethics and design into discourses about ethics and design",[11] a routinized checklist approach to ethics that is powered by capitalist logics and a technosolutionist mindset.[13] Companies are "learning to speak and perform ethics rather than make the structural changes necessary to achieve the social values underpinning the ethical fault lines that exist".[13] For Greene, Hoffmann, and Stark, these practices are both too focused on technical tweaks, blinded by technical concerns about how to embed fairness and accountability within machine learning systems and neglectful of structural injustice, and are universalist projects "justified by reference to a hazy biological essentialism".[11] For human rights experts such as Paul Nemitz[12] and Phillip Alston who jokingly said at a 2018 AI Now conference that he wanted to "strangle ethics",[13] technology ethics is seen as a substitute or an alternative to more adequate human rights laws.[16]

As argued further below, these critiques ought to be taken seriously. They shed light on the politics of AI and on crucial blind spots that are performatively and voluntarily obscured by corporate ethics practices. Yet they are at their weakest when, instead of understanding that legal and technological governance are necessarily embedded in ethical and moral thinking, they draw sharp dichotomies between "ethics" and "law", between "ethics" and "justice", as if these were incompatible alternatives and they often misconstrue the relation between "ethics" and "politics" failing to take them as all ingredients playing complementary roles in a desirable understanding of technology governance. The author calls ethics bashing the reduction and dismissal of ethics as a simplistic alternative to law or justice, and the lazy conflation of moral thinking and inquiry with a politics of neutral thinking and with appearances of "ethics" that are hardly in line with what morality requires. The author identifies three fallacies that characterize ethics bashing positions.

First, Nemitz has drawn sharp distinctions between ethics and law as separable and discrete practices: the key question, writes Nemitz, is "which of the challenges of AI can be safely and with good conscience left to ethics, and which challenges of AI need to be addressed by rules which are enforceable and based on democratic process, thus laws".[12] Such distinctions operate on the positivist assumption that law—its making, interpretation, and application—are institutional facts whose existence and relevance are entirely distinct and separable from its societal and moral implications. Positivists, frequently relying on a Humean separation of "is" and "ought", or fact and value, argue that law belongs to the realm of positive facts while morality is completely distinct and belongs to the realm of moral value and of the "ought".[72] An understanding of law as conceptually separate from morality obscures how law is constructed—written, interpreted, and applied—in ways that embed certain moral and political commitments. As Dworkin understood and theorized, law has no factual existence other than the existence we give it through the principled moral and political commitments we express as we interpret and apply it.[24] Consequently, the task of understanding, applying, and re-making law is inseparable from engagement in the internal reflexive exercise of moral commitment and ethical evaluation. Instead of saying that law is superior to ethics, we might want to respond to obtuse corporate ethics efforts by saying that a capacious understanding of morality and ethics is incompatible with ethics washing and extensive self-regulation and that morality instead requires effective laws and robust external checks and accountability mechanisms on machine learning systems, especially when they affect vulnerable populations.[73]

The second and third fallacies, the conflation of "ethics" and "self-interested politics" and the distinction between "ethics" and "social justice", are connected. Both attitudes are grounded in a relatively narrow understanding of moral inquiry as a discrete, individual-centric, and particularized exercise whose politics and impact lie in its separateness from broader political and institutional questions. As described by Metcalf et al., the distinction between narrow "ethics" and capacious "justice" became a central focus of discussions during the 2019 ACM Conference on Fairness, Accountability and Transparency.[13] However, justice and morality are inseparably intertwined. Critics are right to argue that the focus on design and on embedding fairness in machine learning is too narrow to address more urgent questions around these technical systems' political dimensions and effects on structural inequality, capitalist exploitation, surveillance, disinformation, and environmental degradation.[10, 13, 14] However, responding to narrow and techno-solutionist corporate approaches on "ethics" is not exhaustively done by arguing somewhat simplistically that justice is superior to ethics, whatever that means, or that ethics has a flawed politics. It must be done by showing that any meaningful understanding of ethics (or politics) must include concerns about structural inequality, capitalist extraction, and environmental justice, or else it is an empty exercise that has little to do with the ethics, justice, and politics of new technologies and their societal impacts.

The answer to instrumentalized ethics is not to draw simplistic dichotomies, but to provide a richer account of how ethics, politics, and law are connected and can work together to enable a better understanding of AI's shortcomings and to foster political and other change. By addressing ethics from the outside, as a discrete practice that does not include them, critics of corporate ethics often fail to recognize that ethics is something they also engage in and that existing corporate practices are in fact morally flawed. The task is therefore to change the way we collectively engage in moral inquiry, equipping ourselves with a better understanding of injustice, inequality, and other digital harms. Corporate logics of profit, expanding production, capitalist exploitation, and so on are often incompatible with a capacious view of morality.

In the remainder of this article, the author articulates what the role of moral philosophy should be in technology policy debates and how a view that takes the reflexive internal exercise of moral inquiry as valuable can shed light on the "ethics washing" debate. The author then concludes with what ethics in technology must look like going forward.

7 The Moral Limits of Corporate Ethics and Self-Regulation

Equipped with a richer understanding of what ethics and moral philosophy are and can do, the question now is what role moral philosophy can play in informing technology policy and particularly the question of what makes ethics-based efforts as practiced in corporate tech settings particularly problematic from a moral philosophical perspective. Moral philosophy can provide a lens to evaluate the moral wrongness of some of these efforts.

As described above, companies such as Google, Apple, Microsoft, OpenAI, Palantir, and Facebook are increasingly making efforts to consider an ethical standpoint. The intentions behind their proactive efforts are often presented as good, but the practices remain driven by market incentives and techno-centric perspectives and motivated primarily by the need to avoid financial and other company risk.[11, 13] Notwithstanding good intentions, therefore, embedding philosophers or ethicists within technology companies appears to be a façade that is frequently used to legitimate certain pre-existing practices and to shield companies from measures more protective of consumers. This is true of corporate settings but also of public institutions. Taylor and Dencik for example have described the political dynamics within the European Commission's High Level Expert Group on AI, showing that instead of having outcomes guided by processes of reflection and philosophical principles, ethical reflections are often designed to produce pre-determined instrumental outcomes.[18] They state that after months of discussion around "red lines" on the use of AI, corporate participants in the High Level Group stated: "the word 'red lines' cannot be in this document … at any point … and the word 'non-negotiable' has to be out of this document."[18] As Taylor and Dencik point out, "if the possibility of delineating meaningful boundaries for technology … is off the table, then so is an important part of the task of ethics."[18]

As we assess these ethics initiatives, we are therefore pulled in two directions. On one hand, we are tempted to welcome some of these developments as positive. On the other hand, we are moved to criticize these efforts for the opportunism they represent. Where we stand on this spectrum will often be informed by our situated perspective, our training, by who pays us, etc. What moral philosophy as a method enables us to do is to take a step back, to consider these attitudes along a spectrum of nuanced positions on companies' ethical behavior, and to evaluate our reasons for supporting or resisting initiatives such as a corporate ethics council or an AI Panel of Experts at EU level. It allows us to suspend our intuitive reactions and take a less polarized perspective on the question: What is wrong with the instrumentalization of ethics language? And what is wrong with ethics boards and self-regulation?

As seen, much of the debate has centered on ethics as a self-regulatory modality of governance and an alternative to law and government regulation. As Javier Ruiz is reported to have stated, "a lot of the data ethics debate is really about how … we avoid regulation. It is about saying this is too complex, regulation cannot capture it, we cannot just tell people what to do because we do not really know the detail."[18] Self-regulation and self-publicity at first both seem benign. Self-regulation in certain cases is not only tolerable but actually welcome, for instance where regulatory interference by a public agency is unlikely to be effective and where a self-regulatory approach can lead to substantive policy improvements for individuals and society. Further, in principle it does not seem morally objectionable to fund and develop initiatives that foster a positive image of one's business, nor does it seem wrong for a business to engage in self-publicity and self-advocacy. However, when looking further the reality is more complex.

To use an example, let us focus on the case of self-regulation in relation to online content moderation on Facebook. In the United States, governmental regulation of online speech is seen with suspicion.[74, 75] The solution to the regulation of online speech on Facebook has consequently materialized in the form of an internal Facebook Oversight Board (FOB), a quasi-judicial body set-up internally but composed of external experts to adjudicate on the acceptability of controversial user content on the platform.[76] The body has been praised as "one of the most ambitious constitution-making projects of the modern era",[77] and is seen as a workable and promising approach for taming Facebook's power over online content in the face of First Amendment restrictions on government regulation.[78] Nonetheless, while the Board may bring about needed transparency and an appearance that content moderation is being tackled fairly, we must look beyond Facebook's messaging to find its shortcomings, procedural and otherwise. In spite of its carefully crafted set-up and the well-intentioned messaging around its existence, it is likely that the FOB will serve the interests of Facebook more than those of users. First, it provides a way to shield Facebook from other forms of regulation and scrutiny on matters of content moderation and community guidelines, including the intervention of national or international courts but also the formulation and enforcement of legislative redlines and constraints. Second, by centering attention on content moderation and community guidelines, it allows Facebook to continue developing its News Feed algorithms as it pleases, and to continue showing individuals lucrative content, without interference from regulators or courts. Thus, far from addressing all questions of online speech harms, the FOB seems to divert attention toward some issues and away from the most pressing concerns around misinformation and political propaganda.[79]

The case of facial recognition technologies is analogous. In the United States, much state regulation of private technology firms is made difficult by the First Amendment.[80] The solution to making facial recognition more ethical was thus for some time believed to be something that must originate within the proprietary walls of tech companies and not something that can be initiated by government entities or the Federal Trade Commission (FTC). But things are changing. Following activist efforts, companies like IBM, Amazon, and Microsoft have scaled back on their offering of general purpose facial recognition software.[81, 82] More recently Facebook has declared that it will cease to use facial recognition.[83] Earlier, company ethics boards themselves, such as Axon's, recognized the importance of public oversight on these technologies.[84] In spite of litigation by tech companies to defend their self-regulatory immunities, it seems that the nomination of Alvaro Bedoya to the FTC will mark a turning point in the relationship between state power and self-regulatory power in this space.

Self-regulatory and ethics washing initiatives such as the FOB, Google's ATEAC Board or Axon's Report on facial recognition technologies should prompt us to look beyond appearances and ask whether their very existence, in spite of appearing useful and a step forward, might in fact performatively obscure more pressing problems and risk long-term harm.

8 A Critique of Ethics Washing from Within Moral Philosophy

To explore the moral limits of these internal corporate efforts superficially aimed at developing more ethical artificial intelligence, we must again turn to moral philosophy. At least three moral arguments can be raised against initiatives that co-opt ethics language and self-regulation for selfish corporate purposes that include profits and reputation.

First, the type of ethics work carried out within companies or ethics boards more often than not seems to lack instrumental value: it does not have beneficial effects on individuals and society, because it is undertaken under conditions that deny these beneficial effects. Second, these practices also seem to lack much of the intrinsic, or independent, value associated with philosophical inquiry insofar as they do not seem to be undertaken in ways that value the process itself and with the aim of achieving overall justice. Third, even if these ethics-based practices were carried out in absolute good faith and in pursuit of justice, and thus maintained both their instrumental and intrinsic value, instrumentalizing ethics reasoning and language to reach company goals entails a specific kind of epistemic concern. Indeed, it seems that the performative role of ethics language remains problematic even where, as the cases of the Facebook Oversight Board or the Axon Ethics Board have illustrated, these efforts are intended to address real issues and in fact could have positive effects. This happens where, in spite of having some instrumental value, these efforts instrumentalize ethics for the sake of other selfish or less valuable ends yet are presented as panaceas that serve the public interest. In what follows I explore these three arguments.

The first critique of self-regulation and company ethics is an argument grounded in the poor instrumental value, or small positive impact, of ethical work performed within a company. Ethics bodies or in-house philosophers are purportedly set up and hired to make a difference to a company's social impact. Yet as long as philosophical inquiry is mandated and funded by a company, and carried out within closed corporate proprietary walls, its primary function is to benefit companies and fulfill their pre-existing mandates, and cannot be to benefit society at large and lead to social renewal. Internal AI ethics practices are frequently put in place for compliance purposes, to pre-empt reputational and financial risk.[13] They are subjected to internal limits, subordinated to the endorsement of high management, and dependent on company funding. This dependency on the company's control renders ethics rhetoric inadequate for addressing serious cases of company misconduct and also unfit for achieving societal change.

The narrow impact of ethics-based efforts carried out within tech companies is due in part to formal limitations on employee-philosophers' or ethics boards' mandates and in part to more diffuse pressures that companies exert on technological discourse and context. Formally, for example, Apple's philosopher in residence Joshua Cohen has been forbidden from making public appearances since he started working for the company and Microsoft's AI ethics board does not disclose the reasons for its decisions.[85] The firing of former Google employees Timnit Gebru and Margaret Mitchell for writing allegedly controversial papers and pushing for a prosocial AI agenda inside the company illustrates companies' power to formally police internal ethics efforts.[6, 7] It also however shows the potentially strong instrumental value of social media backlash following these episodes.[4] Less visibly, companies also exert diffuse influence on the broader discourse around technological innovation and ethics by funding research and policy initiatives that favor their agendas and selecting people to engage with (and whose ideas to highlight), including the people these companies choose to have as part of their ethics-based initiatives.[68, 86]

These internal pressures in turn shape the substance and conservative nature of resulting ethics-based work. Strong pushes for data protection guarantees, data minimization mandates, redlines on the use of AI in credit scoring, policing, criminal procedure, or antitrust enforcement can hardly be initiated by a company's ethics board or in-house philosopher. Their role remains confined to steering, reviewing, and advising on policies and product launches within the confines of existing business models, so as to preserve those business models. For example, in June 2020, IBM publicly announced it would stop offering general purpose facial recognition or analysis software.[81] This move, which was a significant departure from IBM's long-standing position on facial recognition and was followed by similar announcements by Amazon and Microsoft, came as a result of external political pressures in the wake of George Floyd's death in Minneapolis, not as a result of the company's internal ethical compliance processes.[82]

Yet it is precisely at moments of political and moral breakdown, where a company's activities and general goals clearly come into conflict with the interests of society, that ethics can acquire central importance[13] and can provide a fruitful lens for evaluating and deciding the way forward. In most cases, instead, the breakthrough potential of ethics as a mechanism for learning from and facing dilemmas and contradictions is missed. As long as the ultimate decision-maker on any given AI policy is the company itself, as long as internal ethics programs are focused on rhetoric more than on substance, these initiatives will keep benefiting the industry more than users and their instrumental value for society is limited.

The second critique of so-called ethics washing looks at the act of engaging in these efforts by philosophers-in-residence, or members of ethics boards, and examines the intrinsic or independent value of these people's engagement in moral thinking. Moral philosophy as a practice has value when followed in pursuit of independently valuable goals such as truth, justice, or the well-being of society. To be intrinsically valuable, engaging in moral argument must be done to a substantial extent out of commitment to moral principle, in the belief that it can lead to a better understanding of moral questions. If instead it is undertaken for the sake of earning money, pleasing employers, or obtaining honors and recognitions, it loses some of its special worth.

We might think that this critique is about the actual motivations of the philosophers and experts that engage in the exercise. When looking at cases of philosophers-in-residence, ethics boards, or academics who work closely with these companies, there are doubtless some individuals who do it to raise their profile or create connections that can lead to further work in the field, or even to obtain promotions, honors, or greater impact and salience for their work. Yet many also do it simply because they believe that their involvement might lead to a positive overall impact or in the hope of getting insights into how the company works. It is tempting to focus on these people's intentions and blame their shortsighted mindsets, but focusing on intentions seems unhelpful: the road to hell is paved with good intentions.

To better characterize the independent value of ethics-based work, we must look beyond intentions and instead at scope: actual commitment to moral principle requires questioning what an employer requires. Philosophical thinking must have the potential to reach beyond the limits imposed by companies in corporate settings. For example, saying that a facial recognition algorithm should be reviewed because it systematically identifies white people more accurately than black people seems right but is not sufficient. Rectifying bias requires more than acknowledging that the algorithm needs "fixing". It requires making sure that the algorithm is not deployed in settings where it might cause irreparable harm to black people. It also possibly involves thinking about preventing the use of such algorithms by the police, or by society at large, and replacing them with human decision-making.[10, 56] To the extent an ethics board or in-house philosopher engages in moral argument with a view to correcting the algorithm yet is prevented from considering or voluntarily ignores these other considerations, their moral inquiry seems to lack substantive independent value. Philosophical inquiry achieves its full potential only when it comes with full and unrestricted substantive commitment to moral principle and justice.

Third and finally, even if these efforts did have intrinsic and/or instrumental value, the expression “ethics washing” denotes a particular epistemic function of the activities in question which requires distinct analysis. Ethics rhetoric, as it is funded and constructed in academic and corporate circles, may have the effect of freezing popular imagination and of preventing the emergence of valuable alternatives.[68] It may promote and reinforce a narrow and confined vision of the possibilities for regulatory and societal change.

It can, for example, mislead the public into believing that previously contested policies have now become acceptable, thus creating a legitimacy buffer for objectionable corporate action. Immunizing corporate action from public scrutiny is dangerous for more than one reason: apathy strengthens corporations and weakens activists, it shifts the burden of policing new technologies from deep-pocketed security and defense departments and private companies to poorly funded activist groups and other marginalized stakeholders. It can also discredit awareness-enhancing efforts and narrow the spectrum of contestation and debate. Self-regulatory efforts, such as the example of the FOB provided above, tend to narrow the scope of a debate, marginalizing questions of structural injustice or disruptive change and instead centering attention on procedural fairness and fixable tweaks. This—predictably—ends up favoring incumbents. Although the performative dimensions of ethics washing are hardly visible by a majority of consumers, they are in fact crucial to a comprehensive analysis of corporate and governmental stakeholders’ strategies in this space and of the moral value and acceptability of their efforts.

Overall, an analysis from the perspective of moral philosophy confirms the view of many critics of ethics washing efforts. It helps us see many of these in-house ethics initiatives as lacking significant instrumental and intrinsic value and also as playing a performative function that can negatively affect persons. There are no doubt exceptions of companies really working to ensure that internal ethical work is independent and valuably contributes to a more just society. However, in general policymakers should not overlook the salience and weight of these critiques of ethics as a self-interested rhetoric. Many existing internal efforts to construct a corporate ethics, particularly around AI, largely remain a façade.

9 Avoiding Ethics Bashing

If the reasons for criticizing and resisting ethics washing are ones found within moral philosophy, where does this leave us on the role of moral philosophy? How should we understand corporate ethics? Two main fallacies seem at play in overbroad critiques of ethics that see ethics as distinct from law, politics, justice or social organizing: a linguistic misunderstanding, that is to say the conflation of instrumentalized ethics washing efforts with moral philosophy as a reflexive exercise, and ignorance of or resistance to the possibilities and importance of moral philosophy as a discipline and method.

The linguistic misunderstanding is due to what the author has described above as companies’ cooptation of the language and performative function of “ethics” to pursue self-promotional goals. Instrumentalized and emptied of its instrumental and intrinsic value, what remains of “ethics” is an empty construct trapped between meanings and signifying timid instances of self-regulation, static and finite lists of guiding principles, and other forms of narrow and conservative regulative “fixes”. None of these embodied instances of the practice of ethics are actually likely to be fully morally defensible, but as the word quickly gains traction, it gets defended or criticized at face value by corporations and critics alike. These dynamics further entrench the misuse and instrumentalization of ethics language. In policy circles, the word becomes a red herring, a mode of governance or a communications strategy to dismiss. Yet the misunderstanding at bottom is this: what is called “ethics” may have nothing “ethical” in it. It may have no intrinsic value for those who perform it and may have instrumental value only for those who commission it and not for society at large.

Much of the ink used to bash “ethics” was perhaps justified but it could have been used more wisely by distinguishing corporate ethics, or ethics washing, from the practice of moral philosophy. We too frequently neglect that “ethics” can and must encompass more than what companies make of it: that properly contextualized, ethics can be a valuable methodology for rethinking the competing or complementary merits of different kinds of regulation, including self-regulation and other forms of law and policy-making.

A richer critique of corporate self-regulatory efforts therefore demands that we operate at two levels: be critical of ethics washing, while also being aware that our very critique positions ourselves distinctly within moral philosophy. In other words, when criticizing certain practices we necessarily adopt a distinct moral stance that is within moral philosophy—not outside of it. We must thus be ready to engage more thoroughly with the flaws of narrow approaches to ethics and to accept that defending more capacious ethical stances is related to a better understanding and awareness of moral philosophy’s potential—not a blank rejection of it as a language, practice, discipline, and mode of inquiry. This requires a deep societal reckoning with the values and limits of moral philosophy.

To change tech ethics, it is urgent to rethink the way technology ethics comes to exist and is talked about. Since ethics washing is broadly antithetic to meaningful and capacious ethics, it is important for policy change to originate primarily outside formal and informal corporate settings. To be effective, the role of philosophers, boards, and other formalized bodies concerned to bring about ethical AI must be re-imagined, their scope of action and mandate must extend outside the corporate walls of companies such as Google or IBM, they cannot be exclusively or primarily funded by companies such as Facebook or Palantir, they must to the extent possible safeguard themselves from opportunistic corporate discourse around “ethical AI”. A deep reinvention of the structures, processes and modes of governance through which technological impacts on society are evaluated is urgent. At their core, these processes must facilitate the moral evaluation, questioning, and constant re-assessment of technological developments. Far from treating technological developments as moments of ethical breakdown, technology as a whole must be seen as a system that endemically tends toward societal breakdown, and therefore requires constant reflexive reconsideration, revision, and re-imagination.

Criticized as complex, abstract, apolitical, and misleadingly neutral or objective, philosophy is frequently dismissed in areas such as technology policy which are fast moving, full of ideological conflicts, and in need of quick and effective responses. However, it is clear that quick and effective fixes are not the answer. Ideological conflicts and the pace of innovation are not barriers to doing more impactful and valuable philosophical work in this sector. Indeed, the current technological zeitgeist of strong resistance to surveillance capitalism; new data privacy laws; the complicated relationship between big tech, big oil, and climate justice; tech employee movements and whistleblowing; COVID-19 and Black Lives Matter suggests that something within technology is changing, and that it is time we adopt new tools and modes of thinking to fight technological injustice. What the tech ecosystem is in greatest need of today, in fact, seems to be a slower, richer, more comprehensive investigation of what various technology companies and stakeholders owe to humans, to animals, and to the planet. New technologies are also making us reinvestigate and question the commitments we humans owe to each other, as well as to other beings and to the global planet ecosystem. This is precisely what moral philosophy is for. We may want to stop bashing it and instead invest in re-imagining it.

10 Conclusion

This article has argued that ethics washing and ethics bashing are both reductive tendencies that rely on a limited understanding of what ethics actually entails. Ethical reasoning or moral inquiry can have intrinsic value as a process and instrumental value as a means to the achievement of other valuable outcomes. The author has argued that the more ethics is used in tech circles as a performative façade and the more it is instrumentalized and voided of its intrinsic reflexive value, the less value ethics can have overall as a practice and mode of inquiry. Adopting a perspective internal to moral philosophy helps us see the limits and actual similarities of what seem like polar opposites—ethics washing and ethics bashing—as two instances of instrumentalized ethics language.

The way to combat ethics washing, therefore, is not to instrumentalize, reduce, and then dispose of ethical language, but rather to distinguish performative and instrumentalized forms of ethics from valuable commitments to moral principle that promote advancements in self-knowledge, understanding, and social change. Although philosophers might never fully adapt their methodology to fast-paced and politicized technology environments, we cannot disregard the immense depth and richness that philosophy can bring to any debate, not least ones about technology governance.

We all ask moral questions as part of our daily pursuits. Technology scholars and policymakers should embrace moral philosophy and value its porous, principled, and open-ended richness, yet resist its instrumentalization or reduction to a performative ethics. Moral philosophers should take on the difficult task of rethinking how new technologies interact with humans so as to provide answers to questions in urgent need of theorization. We all ask moral questions as part of our daily pursuits. To avoid falling into reductive epistemic and ideological traps, it is everyone's duty to nourish curiosity for ethics' and moral philosophy's role in tech and beyond. However, before we can re-center attention on technology ethics, value it in our daily pursuits, and renew interest in the interconnections between moral philosophy, justice, politics, and law, it is urgent to de-center the structures for engaging in theoretical and ethical thinking from corporate settings. Making a commitment to moral principle in technology is impossible without a new governance framework that ensures that ethics in technology remains independent and capacious.

Acknowledgment

E. Bietti thanks Jeff Behrends, Yochai Benkler, Brian Berkey, Reuben Binns, Mark Budolfson, Urs Gasser, Ben Green, Lily Hu, Lucas Stanczyk, Luke Stark, Jonathan Zittrain, and some anonymous reviewers for their valuable input on this article.

References

[1] K. Walker, An external advisory council to help advance the responsible development of AI, https://blog.google/technology/ai/external-advisory-council-help-advance-responsible-development-ai/, 2019.

[2] Googlers Against Transphobia and Hate, Google must remove Kay Coles James from its Advanced Technology External Advisory Council (ATEAC), https://medium.com/@against.transphobia/googlers-against-transphobia-and-hate-b1b0a5dbf76, 2019.

[3] S. Levin, Google scraps AI ethics council after backlash: "Back to the drawing board", https://www.theguardian.com/technology/2019/apr/04/google-ai-ethics-council-backlash, 2019.

[4] T. Gebru, I was fired by @JeffDean for my email to Brain women and Allies. My corp account has been cutoff. So I've been immediately fired, https://twitter.com/timnitGebru/status/1334352694664957952, 2020.

[5] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, On the dangers of stochastic parrots: Can language models be too big? in *Proc. the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, Virtual Event, Canada, 2021, pp. 610–623.

[6] C. Newton, The withering email that got an ethical AI researcher fired at Google, https://www.platformer.news/p/the-withering-email-that-got-an-ethical, 2020.

[7] J. Vincent, Google is poisoning its reputation with AI researchershttps://www.theverge.com/2021/4/13/22370158/google-ai-ethics-timnit-gebru-margaret-mitchell-firing-reputation, 2021.

[8] B. Wagner, Ethics as an escape from regulation: From ethics-washing to ethics shopping? in *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*, E. Bayamlioğlu, I. baraliuc, L. Janssens, and M. Hildebrandt, eds. Amsterdam, the Netherlands: Amsterdam University Press, 2018, pp. 84–89.

[9] T. Harris, http://www.tristanharris.com/, 2021.

[10] J. Powles and H. Nissenbaum, The seductive diversion of 'solving' bias in artificial intelligence, https://medium.com/s/story/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53, 2018, .

[11] D. Greene, A. L. Hoffmann, and L. Stark, Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning, in *Proc. of the 52nd Hawaii International Conference on System Sciences*, Honolulu, HI, USA, 2019, pp. 2122–2131.

[12] P. Nemitz, Constitutional democracy and technology in the age of artificial intelligence, *Philosophical Transactions of the Royal Society A*, vol. 376, no. 2133, p. 20180089, 2018.

[13] J. Metcalf, E. Moss, and D. Boyd, Owning ethics: Corporate logics, Sillicon Valley, and the institutionalization of ethics, *Social Research*: *An International Quarterly*, vol. 82, no. 2, pp. 449–476, 2019.

[14] B. Green, Data science as political action: Grounding data science in a politics of justice, *Journal of Social Computing*, doi:10.23919/JSC.2021.0029.

[15] H. L. A. Hart, *The Concept of Law*, Oxford, UK: Clarendon Press, 1961.

[16] L. Hu, Tech ethics: Speaking ethics to power, or power speaking ethics? *Journal of Social Computing*, doi: 10.23919/JSC.2021.0033.

[17] J. E. McNealy, Framing and the language of ethics: Technology, persuasion, and cultural context, *Journal of Social Computing*, doi: 10.23919/JSC.2021.0027.

[18] L. Taylor and L. Dencik, Constructing commercial data ethics, 2020, https://doi.org/10.26116/techreg.2020.001, 2020.

[19] A. L. Hoffmann, Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse, *Information Communication and Society*, vol. 22, no. 7, pp. 900–915, 2019.

[20] C. Grannan, What's the difference between morality and ethics? *Encyclopedia Britannica*, https://www.

britannica.com/story/whats-the-difference-between-morality-and-ethics, 2021.

[21] D. D. Runes, *The Dictionary of Philosophy*. New York, NY, USA: Philosophical Library, 1983.

[22] T. Scanlon, *What We Owe to Each Other*. Cambridge, MA, USA: Harvard University Press, 1998.

[23] R. Dworkin, *Justice for Hedgehogs*. Cambridge, MA, USA: Belknap Press, 2011.

[24] R. Dworkin, *Law's Empire*. Cambridge, MA, USA: Belknap Press, 1986.

[25] O. Tene and J. Polonetsky, A theory of creepy: Technology, privacy, and shifting social norms, *Yale Journal of Law, and Technology*, vol. 16, no. 1, p. 2, 2014.

[26] J. Rawls, *A Theory of Justice*. Cambridge, MA, USA: Harvard University Press, 1971.

[27] J. Cobbe and E. Bietti, Rethinking digital platforms for the post-COVID-19 era, https://www.cigionline.org/articles/rethinking-digital-platforms-post-covid-19-era, 2020.

[28] J. Cheung, Real estate politik: Democracy and the financialization of social networks, *Journal of Social Computing*, doi: 10.23919/JSC.2021.0030.

[29] C. Pateman and C. Mills, *Contract and Domination*. Malden, MA, USA: Polity Press, 2007.

[30] B. J. Grosz, D. G. Grant, K. Vredenburgh, J. Behrends, L. Hu, A. Simmons, and J. Waldo, Embedded EthiCS: Integrating ethics across CS Education, *Communications of the ACM*, vol. 62, no. 8, pp. 54–61, 2019.

[31] C. Fiesler, N. Garrett, and N. Beard, What do we teach when we teach tech ethics?: A syllabi analysis, in *Proc. the 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20)*, Portland, OR, USA, 2020, pp. 289–295.

[32] R. Reich, M. Sahami, J. M. Weinstein, and H. Cohen, Teaching computer ethics: A deeply multidisciplinary approach, in *Proc. the 51st ACM Technical Symposium on Computer Science Education*, Portland, OR, USA, 2020, pp. 296–302.

[33] R. Ferreira and M. Y. Vardi, Deep tech ethics: An approach to teaching social justice in computer science, in *Proc. the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE '21)*, Virtual Event, USA, 2021, pp. 1041–1047.

[34] N. Fraser, Rethinking the public sphere: A contribution to the critique of actually existing democracy, *Social Text*, no. 25/26, pp. 56–80, 1990.

[35] M. Hildebrandt, Closure: On ethics, code and law, in *Law for Computer Scientists and Other Folk*, M. Hildebrandt, ed. Cambridge, MA, USA: Oxford University Press, 2020, pp. 283–318.

[36] J. Taplin, *Move Fast and Break Things: How Facebook, Google, and Amazon Cornered Culture and Undermined Democracy*. New York, NY, USA: Little, Brown and Company, 2017.

[37] P. Foot, The problem of abortion and the doctrine of the double effect, *Oxford Review*, vol. 5, pp. 5–15, 1967.

[38] E. Awad, S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J-F. Bonnefon, and I. Rahwan, The moral machine experiment, *Nature*, vol. 563, no. 7729, pp. 59–64, 2018.

[39] E. Awad, S. Dsouza, A. Shariff, J. -F. Bonnefon, and I. Rahwan, Crowdsourcing moral machines, *Communications of the ACM*, vol. 63, no. 3, pp. 48–55, 2020.

[40] A. E. Jaques, Why the moral machine is a monster, https://robots.law.miami.edu/2019/wp-content/uploads/2019/03/MoralMachineMonster.pdf, 2019.

[41] J. Basl and J. Behrends, Why everyone has it wrong about the ethics of autonomous vehicles, in *Frontiers of Engineering Reports on Leading-Edge Engineering from the 2019 Symposium*, National Academy of Engineering, ed. Washington, DC, USA: The National Academies Press, 2020, pp. 75–82.

[42] J. Fjeld, N. Achten, H. Hilligoss, A. C. Nagy, and M. Srikumar, Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI, *Berkman Klein Center Research Publication No. 2020-1*, doi: http//dx.doi.org/10.2139/ssrn.3518482 .

[43] B. Green, The contestation of tech ethics: A sociotechnical approach to technology ethics in practice, *Journal of Social Computing*, doi: 10.23919/JSC.2021.0018.

[44] B. Mittelstadt, Principles alone cannot guarantee ethical AI, *Nature Machine Intelligence*, vol. 1, pp. 501–507, 2019.

[45] H. Nissenbaum, *Privacy in Context*. Stanford, CA, USA: Stanford University Press, 2009.

[46] S. Prideaux, *I Am Dynamite! A Life of Nietzsche*. New York, NY, USA: Tim Duggan Books, 2018.

[47] C. Baumann, Was Hegel an authoritarian thinker? Reading Hegel's *Philosophy of History* on the basis of his metaphysics, *Archiv für Geschichte der Philosophie*, vol. 103, no. 1, pp. 120–147, 2019.

[48] M. Foucault, *Naissance de la Biopolitique: Cours au Collège de France, 1978–1979*. Paris, France: Editions du Seuil, 2004.

[49] M. Rosen, The Marxist critique of morality and the theory of ideology, in *Morality, Reflection and Ideology*, E. Harcourt, ed. Cambridge, MA, USA: Oxford University Press, 2000, pp. 21–43.

[50] J. Zigon, Moral breakdown and the ethical demand: A theoretical framework for an anthropology of moralities, *Anthropological Theory*, vol. 7, no. 2, pp. 131–150, 2007.

[51] J. Habermas, Reconciliation through the public use of reason: Remarks on John Rawls's political liberalism, *The Journal of Philosophy*, vol. 92, no. 3, pp. 109–131, 1995.

[52] K. Dotson, How is this paper philosophy? *Comparative Philosophy*, vol. 3, no. 1, pp. 3–29, 2012.

[53] L. Winner, Do artifacts have politics? *Daedalus*, vol. 109, no. 1, pp. 121–136, 1980.

[54] M. Hildebrandt, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*. Cheltenham, UK: Edward Elgar Publishing Limited, 2015.

[55] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, 2016.

[56] C. O'Neill, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY, USA: Broadway Books, 2017.

[57] S. U. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, NY, USA: NYU Press, 2018.

[58] V. Eubanks, *Automating Inequality: How High-Tech Tools*

Profile, Police and Punish the Poor. New York, NY, USA: St Martin's Press, 2018.

[59] R. Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge, UK: Polity Press, 2019.

[60] S. Constanza-Chock, *Design Justice: Community-Led Practices to Build the Worlds We Need*. Cambridge, MA, USA: MIT Press, 2020.

[61] C. D'Ignazio and L. Klein, *Data Feminism*. Cambridge, MA, USA: MIT Press, 2020.

[62] R. Foroohar, Year in a word: Techlash, https://www.ft.-com/content/76578fba-fca1-11e8-ac00-57a2a826423e, 2018.

[63] K. Shilton, M. Zimmer, C. Fiesler, A. Narayanan, J. Metcalf, M. Bietz, and J. Vitak, We're awake —but we're not at the wheel, https://medium.com/pervade-team/were-awake-but-we-re-not-at-the-wheel-7f0a7193e9d5, 2017.

[64] T. Rees, Why tech companies need philosophers—and how I convinced Google to hire them, https://perma.cc/2967-8H5R, 2019.

[65] Google, Artificial intelligence at Google: Our principles, https://ai.google/principles/, 2020.

[66] IBM, Report: Advancing AI ethics beyond compliance, https://www.ibm.com/thought-leadership/institute-business-value/report/ai-ethics, 2020.

[67] Microsoft, Responsible AI, https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primaryr6, 2020.

[68] R. Ochigame, How big tech manipulates academia to avoid regulation, https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/, 2019.

[69] AI Multiple, AI consulting: In-depth guide with top AI consultants of 2020, https://research.aimultiple.com/ai-consulting/2020, 2020.

[70] Clifford Chance, Tech Group, https://www.cliffordchance.com/hubs/tech-group-hub/tech-group.html, 2020.

[71] L. Floridi, Translating principles into practices of digital ethics: Five risks of being unethical, *Philosophy and Technology*, vol. 32, pp. 185–193, 2019.

[72] D. Hume, *A Treatise of Human Nature*, London, UK: Penguin Classics, 1739.

[73] S. Viljoen, The promise and limits of lawfulness: Inequality, law, and the techlash, *Journal of Social Computing*, doi: 10.23919/JSC.2021.0025.

[74] J. Balkin, Free speech is a triangle, *Colorado Law Review*, vol. 118, p. 201, 2018.

[75] A. Shanor, The new Lochner, *Wisconsin Law Review*, vol. 1, pp. 133–208, 2016.

[76] C. Botero-Marino, J. Greene, M. W. McConnell, and H. Thorning-Schmidt, We are a new board overseeing Facebook. Here's what we'll decide, https://www.nytimes.com/2020/05/06/opinion/facebook-oversight-board.html, 2020.

[77] E. Douek, Facebook's "oversight board:" Move fast with stable infrastructure and humility, *North Carolina Journal of Law and Technology*, vol. 21, no. 1, pp. 1–78, 2019.

[78] T. Kadri and K. Klonick, Facebook v. Sullivan: Building constitutional law for online speech, *Southern California Law Review*, vol. 93, p. 37, 2019.

[79] S. Vaidhyanathan, Facebook and the folly of self-regulation, https://www.wired.com/story/facebook-and-the-folly-of-self-regulation/, 2020.

[80] J. Jaffer and R. Krishnan, Clearview AI's first amendment theory threatens privacy—and free speech, too, https://slate.com/technology/2020/11/clearview-ai-first-amendment-illinois-lawsuit.html, 2020.

[81] IBM, IBM CEO's letter to congress on racial justice reform, https://www.ibm.com/blogs/policy/facial-recognition-susset-racial-justice-reforms/, 2020.

[82] A. Smith, IBM will no longer develop facial recognition technology following George Floyd protests, https://www.independent.co.uk/life-style/gadgets-and-tech/news/ibm-facial-recognition-george-floyd-protests-a9556061.html, 2020.

[83] J. Pesenti, An update on our use of face recognition, https://about.fb.com/news/2021/11/update-on-use-of-face-recognition/, 2021.

[84] Axon, First report of the Axon AI ethics board: Face recognition, https://www.policingproject.org/axon-fr, 2019.

[85] A. Papazolgou, Silicon Valley's secret philosophers should share their work, https://perma.cc/6KZR-ASJ9, 2019.

[86] O. Williams, How big tech funds the debate on AI ethics, https://perma.cc/5999-57BW, 2019.

Elettra Bietti is pursuing the PhD degree at Harvard Law School. She is an incoming joint postdoctoral fellow at NYU Law's Information Law Institute and at the Digital Life Initiative, Cornell Tech. She is affiliated to Harvard's Berkman-Klein Center, Harvard's Weatherhead Center and Yale Law School's Information Society Project. Prior to her doctorate, she was a competition and intellectual property lawyer in London and Brussels, handling corporate transactions and patent disputes. She received the LLB degree in law from University College London, the LLM degree from Harvard Law School, and the professional diploma in intellectual property law and practice from Oxford University in 2011, 2012, and 2016, respectively, and is admitted to practice law in New York, NY, USA and England and Wales, UK.