

# YOUR THOUGHTS FOR A PENNY? CAPITAL, COMPLICITY AND AI ETHICS

CORINNE CATH AND OS KEYES

## Introduction

The rise in concerns about the harms of algorithmic technologies and reformist efforts to address these harms is accompanied by a critical flank of research and researchers who advocate more radical interventions—and for good reason. Concerns about ‘ethics-washing’, ‘ethics-shopping’ and ‘ethics theater’<sup>1</sup> are repeatedly validated by examples of companies seeking to water down research that brings scrutiny to inherent harms rooted in AI business models and corporate power consolidation.<sup>2</sup> These concerning dynamics occur against the backdrop of the growing lobbying power of the tech sector,<sup>3</sup> as well as increased neoliberalization that sees government and independent sources of funding for academic research reduced and supplanted by industry funding. The growing role of such corporate funding with a direct interest in business-friendly results is distinctly visible in academic research on AI ethics.<sup>4</sup>

This worrisome dynamic brings up a number of issues about what it means to do research into the ethics of AI technologies. It raises questions about who ‘owns’ AI ethics,<sup>5</sup> and by extension: who is responsible for maintaining ethical standards in academic research? Similarly, the role of private industry in research funding raises questions about the subtle and pernicious

- 
- 1 See Mona Sloane, ‘Inequality is the Name of the Game: Thoughts on the Emerging Field of Technology, Ethics and Social Justice’, Weizenbaum Conference, Berlin, 16–17 May 2019; also Ben Wagner, ‘Ethics As An Escape From Regulation. From “Ethics-Washing” To Ethics-Shopping?’, in *Being Profiled*, Amsterdam: Amsterdam University Press, 2018; and Corinne Cath and Fieke Jansen, ‘Dutch Comfort: The Limits of AI Governance Through Municipal Registers.’ *arXiv e-prints* (2021): arXiv–2109.
  - 2 Algorithm Watch, ‘AlgorithmWatch Forced to Shut down Instagram Monitoring Project after Threats from Facebook’. *AlgorithmWatch*, 13 August 2021, <https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook/> and Issie Lapowsky, ‘How Tech Giants Court and Crush the People Who Study Them’, Protocol—The People, Power and Politics of Tech, 19 March 2021, <https://www.protocol.com/nyu-facebook-researchers-scraping>.
  - 3 Pawel Popiel, ‘The Tech Lobby: Tracing the Contours of New Media Elite Lobbying Power.’ *Communication Culture & Critique* 11.4 (2018): 566–85 and Max Bank, Felix Duffy, Verena Leyendecker, and Margarida Silva, ‘The Lobby Network; Big Tech’s Web of Influence in the EU’, Corporate Europe Observatory and Lobby Control e.V., 2021, <https://corporeeurope.org/en/2021/08/lobby-network-big-techs-web-influence-eu>.
  - 4 Corinne Cath, ‘Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges’, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376 (2018) and Agathe Balayn and Seda Gürses, ‘If AI Is the Problem, Is Debiasing the Solution?’ Brussels, Belgium: EDRI, 2021, <https://edri.org/our-work/if-ai-is-the-problem-is-debiasing-the-solution/>.
  - 5 Jacob Metcalf and Emanuel Moss, ‘Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics’, *Social Research: An International Quarterly* 86, (no. 2, 2019): 449–76.

ways in which corporations exert pressure over academic work, even when funding is formally described as ‘no strings attached’. Last but not least, the changing landscape of academic funding leads to novel challenges regarding the impact of industry influence on the research agendas, as well as the career trajectories, of the next generation of academics, like the authors of this contribution.

The key question we ask is where and how one draws the line—a question that is both descriptive and normative. In this chapter, we examine both aspects by drawing from auto-ethnographic methods<sup>6</sup> and presenting our experiences in the form of stories in which the authors, each entangled in this reality of industry funding in different ways, reflect on our experiences. We examine how our work is shaped by industry funding, how we negotiate our own lines in the sand regarding when or how we are paid ‘pennies for our thoughts’, and how these negotiations and lines evolve over time. Engaging in both individual reflection and dialogic exchange, we ask ourselves (and each other):

*What lessons can be learned from the ethnographic realities of working on AI in academic settings in which research is reliant on industry funding?*

Answering this question, one that confronts many researchers, will help us provide new insights into how corporate power plays out in the context of academic research on AI ethics.

We argue that although boundaries can be (and often are) drawn between ‘conventional’ and ‘critical’ data ethics perspectives and actors, reality is often more complex. Engagement does not mean an absence of misgivings, as Su, Lazar & Irani adroitly document.<sup>7</sup> Likewise, critique does not avoid complicity.<sup>8</sup> Behind every perspective and position are *people*, navigating their own relation to, and alienation from, their values in a research environment of finite resources and corporate capture.<sup>9</sup>

Even absent such an environment, some degree of compromise is a consequence of having to work with and in relation to other people and organizations with different interests and perspectives. Indeed, as Nick Seaver puts it, one vital site of ethics is how ‘people negotiate among apparently competing values’.<sup>10</sup> Our chapter elucidates what that negotiation can look like in practice and what lessons can be learned for others trying to navigate what it means to do, and relate to, ethics in academic settings and industry-funded AI research.

---

6 Carolyn Ellis, Tony E. Adams, and Arthur P. Bochner, ‘Autoethnography: An Overview’, *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 12.1 (2001) and Deborah Reed-Danahay, *Auto/Ethnography: Rewriting the Self and the Social*, Oxford; New York: Routledge, 1997.

7 Norman Makato Su, Amanda Lazar, and Lilly Irani, ‘Critical Affects: Tech Work Emotions Amidst the Techlash’, *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021).

8 Fred Moten and Stefano Harney, ‘The University and the Undercommons: Seven theses’, *Social Text* 22.2 (2004).

9 Cath, ‘Governing Artificial Intelligence’ and Meredith Whittaker, ‘The Steep Cost of Capture’, *Interactions* 28.6 (2021): 50–5.

10 Nick Seaver ‘CARE AND SCALE: Decorrelative Ethics in Algorithmic Recommendation’, *Cultural Anthropology* 36 (no. 3, 2021): 509–37.

Two themes surface throughout this reflection. First, the power of corporations to indirectly shape research, not by setting the content but rather by drawing the contours of acceptable critique. Second is the relationships that are possible and impossible due to the role of industry funding. The emergence of these two distinct themes and the role of feminist refusal in our shared experience,<sup>11</sup> as PhD researchers who started their research three years and 4,488 miles apart, speaks to the expansive influence of corporate power, the implications this has for research, and the possibilities for the next generation of scholars working on AI ethics through industry funded means.

This chapter is organized as follows. First, we position our work and experiences in recent literature on navigating values in academia, by focusing on the capital, complicity, and articulation work of AI ethics. In doing so, we demonstrate a link between our individual experiences as junior scholars working on AI ethics and the broader dynamics of capture of, and resistance through, research documented in the literature. We show that the work of many scholars working at the intersection of AI and ethics is shaped by the funding realities of contemporary academia, and that efforts to simplistically escape these realities are themselves entangled in questions of power. Second, we introduce two short auto-ethnographic stories (in the sense used by Sotiropoulou and Cranston),<sup>12</sup> based on our research experiences.

Corinne considers the subtle power dynamics of industry funding on AI ethics research and its implications for the next generation of scholars. Os focuses on how funding shapes the network of relations that researchers build and those that, by extension, remain out of reach. We subsequently put our respective experience in conversations with each other, to draw broader conclusions about how AI ethics relates to questions of power, a key theme of this edited collection. We conclude by arguing that it is increasingly difficult for academia to be a bastion of critical thought on AI, given its enmeshment with industry interest and networks. If it ever was possible for the academe to play such a role,<sup>13</sup> these conclusions call for a radical rethinking of how research on AI ethics is funded and the creation of novel networks of counter-power and care, within and outside of the academe.

## Positioning Us: Ongoing Debates About AI Ethics

Efforts to describe and/or demarcate AI ethics often rely on the idea of ‘waves’—of successive, differing lines of scholarship and thought.<sup>14</sup> Although this is common, we worry that it risks rendering invisible and unintelligible scholarship that was ‘out of time’: too radical, too early, or

---

11 Bonnie Honig, *A Feminist Theory of Refusal*, Cambridge: Harvard University Press, 2021 and Audra Simpson, ‘The Ruse of Consent and the Anatomy of “Refusal”: Cases from Indigenous North America and Australia.’ *Postcolonial Studies* 20.1 (2017): 18–33.

12 Panagiota Sotiropoulou and Sophie Cranston, ‘Critical Friendship: An Alternative, ‘Care-ful’ Way to Play the Academic Game.’ *Gender, Place & Culture* (2022).

13 Thao Phan, et al., ‘Economies of Virtue: The Circulation of “Ethics” in Big Tech’, *Science as Culture* (2021): 121–35.

14 Frank Pasquale, ‘Machines Judging Humans: The Promise and Perils of Formalizing Evaluative Criteria’, in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, New York, 7–8 February 2020.

too conventional, too late.<sup>15</sup> Instead, we would subdivide AI ethics research based on its *form of critique*: into internal critique that seeks to alter scholarship from ‘within the [scholarship] itself’ and *immanent critique* (which, while involved with the practices being analyzed, relies on ‘context-transcending claims’).<sup>16</sup>

Internal critique operates within the norms of existing forms of governance. In the case of AI ethics, we can see this in proposals for voluntary or involuntary codes of ethics,<sup>17</sup> proposals to operationalize and orient AI researchers towards notions of ‘algorithmic fairness’, and attempts at developing binding regulations that implicitly further the inevitability thesis of AI—that many politicians, academics, and business leaders subscribe to.<sup>18</sup> It is often this type of critique that finds favor with corporate funders and, as such, has taken up much of the air in discussions about AI ethics.

Operating counter to this, immanent critique draws from broader normative understandings in the hope of achieving not only reformist, but transformative, effects. In the context of AI, this is often referred to as ‘critical data studies’, i.e., research that takes issue not only with the outcomes of algorithmic systems but also with the very premises on which such systems are designed, the political economies that make them possible, and the discursive cultures AI simultaneously depends on and enables.<sup>19</sup> For example, Green and Viljoen outline the limits of algorithmic thought and its technical formalism to argue that many of the harms of AI occur outside of these limits and need to be addressed in social rather than technical terms.<sup>20</sup> Similarly, in a recent report for European Digital Rights organization EDRi,<sup>21</sup> the authors argue that the root causes of power imbalances of AI must be addressed. Likewise, Powles and Nissenbaum argue that deference to AI by focusing on improving the technology from within limits not only critical conversation about the role of technology, but also the role of ethics, law, and the media in limiting AI harms.<sup>22</sup> Critical scholars like Powles, Nissenbaum, Balayn, and Gürses contend that focusing on ‘technical debiasing of systems’, while popular, is insufficient and oversimplifies the socio-technical impacts of AI as functions of its technical design. They make a strong statement that the seduction of such internal critiques and improvements

15 Clare Hemmings, *Why Stories Matter*, Durham: Duke University Press, 2011.

16 Rahel Jaeggi, *Critique of Forms of Life*, Cambridge: Harvard University Press, 2018.

17 See Anna Jobin, Marcello Ienca, and Effy Vayena, ‘The Global Landscape of AI Ethics Guidelines’, *Nature Machine Intelligence* 1.9 (2019).

18 Cath & Jansen, ‘Dutch Comfort’.

19 Lina Dencik, ‘The Datafied Welfare State: A Perspective from the UK’ in Andreas Hepp, Juliane Jarke, and Leif Kramp (eds) *The Ambivalences of Data Power: New Perspectives in Critical Data Studies*, Palgrave Macmillan UK: 2021; Cami Rincón, Os Keyes, and Corinne Cath, ‘Speaking from Experience: Trans/Non-Binary Requirements for Voice-Activated AI’, *Proceedings of the ACM on Human-Computer Interaction* (2021), and Linnet Taylor, ‘Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector’, *Philosophy & Technology*, 2021.

20 Ben Green and Salomé Viljoen, ‘Algorithmic Realism: Expanding the Boundaries of Algorithmic Thought’ in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency 2020*, Barcelona Spain, 2020: ACM.

21 Balayn and Gürses, ‘If AI Is the Problem, Is Debiasing the Solution?’.

22 Powles, Julia, and Helen Nissenbaum, ‘The Seductive Diversion of “Solving” Bias in Artificial Intelligence’, Medium, 2018, <https://medium.com/s/story/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53>.

belies their simplicity, rather than its suitability for addressing the inevitable questions of power and governance invoked by AI.

Understandably, these researchers engaging with the socio-technical nature of AI have turned their eye to ‘AI ethics itself’. Conventional AI ethics, they argue, is hobbled by and fundamentally tied to the economic, social, and political interests of AI industry leaders. This type of immanent critique is less likely to receive industry funding, given its focus on deconstructing the power relations that sustain the hype around AI systems.

Yet there is a growing need for such immanent critique that highlights what AI ethics ‘does’. This type of critique can serve as a meta-analysis of the field of AI ethics and push back against narrow critique that elides structural change to the political and economic realities that sustain it. Researchers worry, for instance, that proposals for ‘fairness’ or ethical codes risk constituting forms of ‘ethics washing’,<sup>23</sup> providing the illusion of ethical behavior over harmful organizations and processes. Similarly, there are many concerns that conventional ideas of AI ethics—and the practitioners engaging in them—are complicit in enabling unjust futures, with the vast inequalities in political and economic power between technology companies and the communities they harm resulting in a field at risk of, if not already subject to, ‘regulatory capture’.<sup>24</sup>

On both sides of this often tense divide are not only principles and practices but also *people*. These people, contrary to classical sociology, are not ‘cultural dopes’ ignorant of the wider processes they participate in and advance. Instead they are agentic creatures often aware of and (as Su et al, and Cath, respectively, demonstrate)<sup>25</sup> responsive to this back-and-forth. The question then becomes what these response could look like; how people already navigate these tensions, and how to do so in a ‘better’ way.

We, the authors, know this all too well, because we are both early-career scholars working in the field of AI ethics. Corinne Cath recently finalized her PhD in the UK and Os Keyes is currently a PhD Candidate in the United States. Before her PhD Corinne worked in politics and for various non-profit organizations; before theirs, Os was a data scientist at both for- and non-profit organizations. Our respective research topics—civil society participation in debates about technology governance and structures of domination—reflect our personal backgrounds, interests, and identities. In acknowledging positionality as queer, trans, white, and European scholars we recognize our backgrounds are irrevocably tied to our research and how we pursue the creation of knowledge about AI ethics. Throughout this piece, we draw from auto-ethnographic epistemologies and methods to outline how knowledge about AI ethics cannot be separated from the people undertaking that research, and how their expertise is (epistemically or materially) supported. The auto-ethnographic stories below are synthesized write-ups of our experiences and conversations that reflect in broad strokes

---

23 Wagner, ‘Ethics as an Escape From Regulation’.

24 Whittaker, ‘The steep cost of capture’.

25 Su et al., ‘Critical Affects’ and Cath, ‘Governing Artificial Intelligence’.

the types of issues we encountered.<sup>26</sup> These broader compilations of our experiences demonstrate the pervasiveness of the influence of industry on AI ethics research across the different contexts and time spans we did our PhD research.

## Corinne's Story: The Impossibility of 'No Strings Attached' Industry Research Funding

In 2016, I started a PhD at a famously old university in Europe. Over the course of my PhD program, I took and taught courses with various professors at this university who included me in ongoing research projects. It was exciting and validating to be involved in their research. Given the proximity of my university to the London-based tech scene and my background as an anthropologist of computing cultures, some of the work I became involved in naturally included research for large, well-known social media companies. Often, this work would be around particular thorny topics, like the ethics of using AI in content moderation at scale. Sometimes this research would be peer-reviewed and public facing, in other cases it would be for internal company purposes. Sometimes the senior researchers I worked with received funding from these tech companies to undertake research.

I had a personal interest in being involved in such industry-driven efforts; it allowed me to peek inside the machine and get a sense of what priorities and politics drive internal decision-making in Big Tech companies. The focused nature of the research, and the trust relationship between the senior researchers I worked with and the companies, meant that the information shared was more elaborate and the people we spoke to less reserved. These experiences provided important insights for an anthropologist of computing communities and cultures, like myself. Even if I would not be able to explicitly write about what I heard or saw, doing this kind of work, I reasoned, would sharpen my overall understanding of the tech sector, and thus my critique. I still believe this to be the case but looking back now, I see how such research can expand the toolbox of these companies in pushing back critics as much as it sharpens researchers' analysis of harms caused.

Notably absent in my industry projects was explicit engagement with some of the most well-known eroding and harmful effects of the tech sector: its surveillance-based data collection and business model. This absence is surprising, given existing academic critique of this business model,<sup>27</sup> and the extent to which it is at the root of many of the ethical concerns that follow. When I made suggestions that would touch on these business models, it was often the senior researchers that would be the first line of defense against including such critiques in our research. The response I received would inevitably read along the lines of:

---

26 Carolyn Ellis, Tony E. Adams, and Arthur P. Bochner, 'Autoethnography: An Overview' *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 12.1 (2011) and Ingo Winkler, 'Doing Autoethnography: Facing Challenges, Taking Choices, Accepting Responsibilities', *Qualitative Inquiry* 24.4 (2018).

27 Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge: Harvard University Press, 2016, and O'Neil, Cathy, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Crown, 2016.

'We can't tell this company to change their business model, even though we all know it is part of the problem, because they will not give up their golden goose.'

On multiple occasions, I inquired whether there was something in the terms of reference of our research grants with the companies that prevented us from highlighting these structural concerns. I would receive many answers that eventually were a version of the following: 'No, we can say what we want. There are no strings attached to this funding. But we will not tell them to change the business model.'

The idiom 'no strings attached' meant that the funding presumably came without any requirements for the content of the research. This statement was true in only the narrowest reading of the sentence. The signed agreements between academics and industry funders might not have spelled out any limitations on the nature of the critique, or its direction towards the structural forces underpinning industry success. Yet, the unspoken agreements around scope, including the practice of doing the research without veering into immanent critique, did pose real limits. In many of the steps preceding the start of the research, industry funders took decisions that meant that they were likely to invite academics that were willing to improve industry status quo, rather than 'disrupt' it.

There is a clear place in academia for people who want to provide internal, rather than immanent, critique by working with companies. I do not critique that choice. The problem, however, is the silent power exerted through funding, in that companies do not mete out funding to academics equally. Rather, various companies, as I learned through being involved in such efforts, seek out academics whose research agendas align with their corporate incentives. Or in other words, they select research that stays within the bounds of internal critique. There are of course exceptions to this rule, and I do not mean to say that all academics who receive industry funding are stopped from making hard-hitting structural critiques. That being said, I saw firsthand how tech companies employ strategic amplification and elevation strategies. A dynamic that Whittaker describes as: 'Industry elevates their weakest critic.'<sup>28</sup>

Industry can and does shape research by creating and strengthening a network of researchers that presents soft, stable, and predictable critique that does not focus too heavily on disrupting the status quo. It is the relations that industry builds with certain academics where much of their power over research agendas resides. This dynamic has direct implications for junior scholars. Industry funding shapes research; not by directly commanding certain outcomes but by amplifying questions that industry is comfortable answering. It thereby shapes the contours of reasonable critique and which harms industry should answer to or not. Simultaneously, funding such internal criticism allows these companies to appear earnestly engaged with academic research while enabling them to sideline those academics deemed too radical, too critical, or too daring.

There are a number of recent examples of this dynamic playing out beyond my direct experiences that bolster my analysis. Recent examples include Google funding a university

---

28 Whittaker, 'The Steep Cost of Capture'.

think tank that champions critique of applying anti-trust regulation to Big Tech on the one hand<sup>29</sup>, and firing internal and external academic critics on the other<sup>30</sup>. We each need to make a decision about our level of comfort with this reality—and root our decisions to accept industry funding in a strong set of explicit politics that allows us to carefully weigh the impact of complicity or refusal, not just for ourselves but for the kind of relations it encourages us to build, with whom and for whom. In their section, Os will pick up on precisely this question of relationship building and how it features in the ability of industry funding to shape research agendas.

## Os's Story: 'All Our Relations', or, How Networking Distorts Networks

In early 2019 I was awarded a Microsoft Ada Lovelace Fellowship. This was very prestigious: of 100 finalists (approximately—I was too busy trying not to throw up from nerves to count), there were five winners, across nearly every STEM field. But it was a big deal regardless of the prestige. It offered three years of tuition, salary, and consequently, three years of focusing on research.

Unlike Corinne's experience, the fellowship didn't come with any ongoing involvement by the sponsor. The expectations were simply that I'd do research that interested me (and stay enrolled). But that lack of interference didn't mean that enabling me was the only consequence. Two others are worth highlighting: the impact the fellowship had on Microsoft, and the impact it had on my personal—professional relationships within research.

The impact on Microsoft made itself known pretty quickly. One of my areas of research expertise is facial recognition (FRT), a technology I firmly believe should be staked through the heart and left at a crossroads. At this point, hating FRT is a family tradition: one grand-advisor was worried about it ten years ago.<sup>31</sup> Another was worried about it *two decades* ago.<sup>32</sup> So it came as an unpleasant surprise—although not, in my most cynical moments, an unexpected one—when I found myself at the Washington state legislature to testify as an expert witness on the need to regulate FRT, where we were opposed by ... Microsoft. Their argument was multifaceted, but one vital part of it was that regulation was unneeded because they could be trusted to address FRT's various biases themselves. Specifically, what FRT needed was the involvement of diverse, expert stakeholders from a range of marginalized communities. In other words: people like me.

---

29 Daisuke Wakabayashi, 'Big Tech Funds a Think Tank Pushing for Fewer Rules. For Big Tech', *The New York Times*, 24 July 2020.

30 Cade Metz and Daisuke Wakabayashi, 'Google Researcher Says She Was Fired Over Paper Highlighting Bias in A.I.', *The New York Times*, 3 December 2020.

31 Susan Leigh Star, 'This Is Not a Boundary Object: Reflections On the Origin of a Concept', *Science, Technology, & Human Values* 35.5 (2010): 601–17.

32 Phil Agre, 'Your Face is Not a Bar Code: Arguments Against Automatic Face Recognition in Public Places', *Whole Earth* 106.2001 (2001): 74–7.

Now: no company is a monolith. I am perfectly willing to accept that a large segment of Microsoft Research (who funded the fellowship) are opposed to FRT, even as the company as a whole advocates and sells it. And, obviously, the lawyer did not specifically mean me; they had no idea who I am. But the fact of the matter is that the *reason* Microsoft can spend its money paying someone who wants to set fire to one of their products is because they feel it does something for them. It lets them pretend the solution is ‘in-house’; it lets them recuperate critique. It positions fellows as in the same role that Fred Moten skewers critical academia for playing; as tasked with *perfecting* technoscience, in being not an alternative to conventional ways of doing but ‘its attempted completion’.<sup>33</sup> This is not to say critical, funded work cannot be effective. It is to say that critique is only *one* of its effects.

Even if the fellowship had not (in a tiny way) changed a Microsoft PR strategy, it certainly changed me. One way in which this happened was via a common secondary clause in the fellowship: that as well as a stipend and tuition, fellows also got priority access to Microsoft Research internships. On the surface, this seems like a win–win; Microsoft gets to try to persuade PhD students of the joys of joining them full time post-graduation, while those students get to eat something other than ramen.

But there’s another consequence, too, of doing the work, of being in that place, with those people. To be in an internship is to be in relation to people, to have a drastically lowered cost of access to some (very smart and kind) researchers. I have met some of my favorite people, including Corinne, through such opportunities. But thanks to the limited time we have on this earth and our constrained capabilities within it, we can only relate to so many people. And so, by creating the possibilities of close relations with some, these opportunities limit the possibility of relations with others. The fellowship afforded me access to new collaborators, and expanded the time available for interactions within professional research circles. But in doing so, it took away from (and weakened) my ties to the people who had got me this fellowship in the first place; to the people and communities that motivate my work.

## Discussion

Conventionally, a discussion section acts to wrap an analysis together, exploring the underlying themes and then compressing them into fixed lessons, lenses, or implications for practice. In attempting to write this, we struggled to point unambiguously in any one direction, for the simple reason that there are, here, no easy answers. Jones states that ‘the choice...[is] between research that is ‘engaged’ or ‘complicit’’,<sup>34</sup> but it is hard to find a spot free of complicity (if such a spot even exists).

---

33 Moten and Harney, ‘The University and the Undercommons’.

34 Richard G. Jones, ‘Putting Privilege Into Practice Through “Intersectional Reflexivity”’: Ruminations, Interventions, and Possibilities’, in *Reflections: Narratives of Professional Helping* (2010), quoting Dwight Conquergood, ‘Between Rigor and Relevance: Rethinking Applied Communication,’ *Applied Communication In the 21st Century*, London: Routledge, 1995.

Rather than write a discussion-section-as-usual, then, we drew inspiration from the conversational approach of Bellanova et al.,<sup>35</sup> and the conversational approach of our writing process. Like Sotiropoulou and Cranston,<sup>36</sup> we see our conversational approach, and the friendship that belies it, as building the radical practice of care ethics in academia. This practice is needed to critically reflect on, and where necessary resist, the influence of industry funding on independent research. We decided to present a series of lenses that are *not* fixed; that are ambiguous, uncertain, sometimes in tension, just like the dynamics we are discussing.

From these lenses, we identified two answers to the question ‘*What lessons can be learned from the ethnographic realities of working on AI in academic settings in which research is reliant on industry funding?*’ The first theme is the importance of the networks and relationships created and never pursued due to the steering force of industry funding. The second theme is the soft and pernicious power of industry funding and its eroding impact on critical research at both the individual and systemic level.

## I: Answering: Just Say No

The most instinctive response to the quandaries of complicity is to demand a ‘no’; to demand the purity of refusing to be complicit at all; to demand that a researcher refuse to stand there. Such a view has the advantage of moral clarity, but quickly becomes messier, precisely as researchers such as Shotwell and scholars of explicitly feminist notions of refusal have made clear.<sup>37</sup>

But who, quite literally, can afford to do this? Research into AI ethics is increasingly precarious outside of the large for-profits we are concerned with, and there is differential access to those resources that *do* exist. Anecdotally, but unsurprisingly, we have also seen differential consequences for saying ‘yes’. The senior academics in Corinne’s example can survive the reputational cost of engaging with such funders—the stock of their power makes it easier for people to excuse their complicity, and harder to experience the consequences, than more junior scholars.

Second: where do we stand otherwise? Universities have their own moral ambiguity, as has already been noted by Phan et al.<sup>38</sup> Shifting to community-based groups and networks is the ideal, but those groups often lack the funding to sustain researchers, and have many more immediately critical things to spend it on. Here, non-industry and government funders could play a critical role to ensure that there are many hubs of organizations and individuals doing critical work. A good recent example is the Distributed Artificial

---

35 Bellanova, Rocco, Kristina Irion, Katja Lindskov Jacobsen, Francesco Ragazzi, Rune Saugmann, and Lucy Suchman. ‘Toward a Critique of Algorithmic Violence.’ *International Political Sociology* (2021.)15 (1): 121–50. <https://doi.org/10.1093/ips/olab003>.

36 Sotiropoulou and Cranston, ‘Critical Friendship’.

37 Alexis Shotwell, *Against Purity: Living Ethically in Compromised Times*, Minneapolis: University of Minnesota Press, 2016.

38 Phan et al., ‘Economies of Virtue’.

Intelligence Research Institute (DAIR) set up by Dr. Timnit Gebru, funded by a number of not-for-profit foundations. More such places of counter-power are needed—as is structuring them in a way that recognizes the risks of associating ‘non-profit’ with ‘beneficial to a community.’<sup>39</sup>

## II: Answering: Take the Money and Run

‘The only ethical relationship with the university,’ Fred Moten says, ‘is a criminal one,’<sup>40</sup> by which he means that researchers should maximize what they take from an institution that demands a ruthless work ethic (and operates from an unjust premise) but offers little resources in return. If that motto is true of universities, it goes double for many companies in the tech sector. The tension of participation—that these organizations do have resources we want—is resolved (or at least, altered), by taking the money and running.

The concern with that approach is that it is a tactic and not a strategy. It might be possible to get an industry-backed grant once or twice and run. Yet eventually, industry funders will catch on. Furthermore, when it comes to supporting research, even fierce industry competitors exchange notes. This should come as no surprise to those monitoring the same five professors being asked to sit on the ethics board, council, or oversight mechanism *du jour*. A take-the-money-and-run reputation will eventually precede you, closing down possible relationships—criminal or otherwise. Additionally, taking your penny and running will eventually lead to a situation where you are functionally saying ‘no’, by being excluded from the pool of fundable academics, and, because of the collective nature of reputations, you may be seen as speaking for your students as well.

## III: Pivoting the Question: Feminist Refusal

Part of the reason these two answers are so dissatisfying is that there is no such thing as a transcendently satisfying answer. Another, however, is that they ask the *wrong question*. In both cases, the question is, ‘What do I, a singular person, singularly do?’ To approach things this way is to miss two vital facets of feminist and/or virtuous practices; that they are often relational, involving assemblages of people, and that (partly as a result of this) they are *practices*. Answers are contingent on circumstance, and the pursuit of them alters those circumstances and in turn demands a reevaluation of our ongoing actions.

The point where we settle, then, is neither on accepting the costs of involvement nor refusing involvement as standalone answers, but instead on what Bonnie Honig frames as *feminist refusal*.<sup>41</sup> To engage in feminist refusal is not to ‘say no’ and be done with it. Rather, it begins with recognising that we are *never* outside relations of power. As such, we have responsibilities of care to those nestled more deeply within these power structures, to treat refusal as a tactic. To refuse purposefully, potentially-temporarily, with an eye to the future.

---

39 Incite! Women of Color Against Violence (eds), *The revolution will not be funded: Beyond the non-profit industrial complex*, Boston: South End Press, 2007.

40 Moten and Harney, ‘The University and the Undercommons’.

41 Honig, ‘A Feminist Theory of Refusal’.

Or: to remain enmeshed on a contingent basis, for a tactical purpose, in relation with the choices of others.

What does this look like in the case of researcher complicity in AI ethics? One hypothetical way it might manifest is by collaborating on a piece of industry-funded research, but with an explicit understanding of what the limits are of the change the company is willing to instigate and making analysis of those boundaries part and parcel of the research. Another would be to engage in conversation when invited by industry, to learn firsthand how they define the problem and the solution—and use that knowledge to sharpen collective attempts at addressing harms. Yet, another would be to push the boundaries of which voices tech companies listen to by including directly impacted communities in the research design. Or to engage as an educator and use a collaboration to outline alternatives to the status quo, with an aim to generate change through internal research.

There are many parts of the research design, from theory, to data to methods, to the accessibility of the findings that should be part of our negotiation process by which we draw research contours and accept industry funding. Feminist refusal, then, is neither inherently incompatible with refusing funding nor embracing it—it simply demands that both be treated as tactics, undertaken with an awareness of, and in relation to, the choices of others. Further, linking to Audra Simpson's parallel thread of indigenous refusal,<sup>42</sup> we would argue that it requires us not only to make choices, but also to examine them on an ongoing basis. To look, in a relational fashion,<sup>43</sup> at the benefits and harms that come from our tentative answers having time to unfold in practice. This reflection needs to happen in and through trusted relationships, or 'critical academic friendship',<sup>44</sup> with peers, students, and mentors. To not simply answer this question but *re-answer* this question, from the perspective of the collective to define what is needed for the collective rather than 'what can I singularly do'.

## Conclusion: What Do Our Experiences Tell us About the State of Research and its Future?

The picture we sketch is complex, it contains multitudes. Yet there are clear lessons to be learned from our experiences as early-career scholars doing research into the ethics of AI technologies. As qualitative researchers, we outline how feminist refusal, friendship, and a reorientation from the individual to the collective provides at least a partial answer to the question of effective resistance. We would also like to offer up reflexivity upon how academic knowledge is created, as key. Part of the problem is that many of the dynamics we described happen in the dark. It is difficult to be open and transparent about industry relations, which in turn makes it difficult to understand exactly how they influence research. We hope that these insights into the nuanced ways in which industry influence permeates academia leads to further develop existing descriptions of industry's impact as well as corporate denial of complicity in dulling tech critique.

---

42 Simpson, 'The Ruse of Consent and the Anatomy of "Refusal"'.

43 Abeba Birhane, 'Algorithmic Injustice: A Relational Ethics Approach', *Patterns* 2.2 (2021).

44 Sotiropoulou and Cranston, 'Critical Friendship'.

Our experiences from the ground up strengthen existing work critiquing the dangers of corporate capture by concretizing how it may influence independent thought in academia. It also provides a clear concern for the future of critical academic research, as few young scholars will be able to escape or avoid the structural pressure put on their institutions to play by the rules, and the machinery, of industry. The deep and resounding consequences of this ‘new normal’ of research raises the question what resistance can and should look like.

We have some thoughts on the question of effective resistance. We call on other early-career scholars to seek out like minded communities and friendships that can support continued critical reflection and dialogue about the impact of individual decisions on the collective body of the academe. Likewise, we ask for tenured faculty to further model and engage in such reflexivity, as they are both closest to the action and least likely to be burned by providing openness. We would like to emphasize the call made by others regarding the need to organize tech workers and academics into collectives that can resist corporate pressure.<sup>45</sup>

As for funders genuinely interested in transformative, rather than reformist, change, we would say: look outside the network. Look at projects and organizations that do not fit the quintessential mold of academic institutions, formal charities and companies. Look at projects that do not promise large-scale payoffs. Look at community groups; look at big changes in small spaces. Accept that your involvement may, in turn, be subject to its own necessary scrutiny. Furthermore, this piece is also a call to action to governments interested in holding tech accountable to seriously consider funding critical academic research. Finding good research will require looking at which scholars and what questions are currently out of bounds for industry funding, and identifying those scholars that are asking research questions from the perspective of the social impact of technology and its collective good, rather than those focused on narrow—penny wise but pound foolish—questions around how to make the AI industry more ethical.

## Funding Disclosures

Corinne’s contribution to this research was not funded; Os is funded by a Microsoft Ada Lovelace Fellowship.

## References

Agre, Phil. ‘Your Face is not a Bar Code: Arguments Against Automatic Face Recognition in Public Places’, *Whole Earth* 106.2001 (2001): 74–77.

Algorithm Watch. ‘AlgorithmWatch Forced to Shut down Instagram Monitoring Project after Threats from Facebook’, *AlgorithmWatch* blog, 13 August 2021, <https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook/>.

Balayn, Agathe, and Gürses, Seda. ‘If AI Is the Problem, Is Debiasing the Solution?’, Brussels Belgium: EDRI, 2021, <https://edri.org/our-work/if-ai-is-the-problem-is-debiasing-the-solution/>.

---

45 See INCITE!, ‘The Revolution will Not be Funded’, Whittaker, ‘The Steep Cost of Capture’, and Ben Tarnoff, *The Making of the Tech Worker Movement*, New York, NY, USA: *Logic* magazine, 2019, <https://logicmag.io/the-making-of-the-tech-worker-movement/>.

- Bank, Max, Duffy, Felix, Leyendecker, Verena, and Silva, Margarida. 'The Lobby Network: Big Tech's Web of Influence In the EU', Corporate Europe Observatory and LobbyControl e.V, 2021, <https://corporateeurope.org/en/2021/08/lobby-network-big-techs-web-influence-eu>.
- Bellanova, Rocco, Kristina Irion, Katja Lindskov Jacobsen, Francesco Ragazzi, Rune Saugmann, and Lucy Suchman. 'Toward a Critique of Algorithmic Violence.' *International Political Sociology* (2021.)15 (1): 121–50. <https://doi.org/10.1093/ips/olab003>.
- Birhane, Abeba. 'Algorithmic Injustice: A Relational Ethics Approach.' *Patterns* 2.2 (2021).
- Cath, Corinne. 'Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376 (2018).
- Cath, Corinne, and Jansen, Fieke. 'Dutch Comfort: The Limits of AI Governance Through Municipal Registers', *arXiv e-prints* (2021): arXiv–2109.
- Conquergood, Dwight. 'Between Rigor and Relevance: Rethinking Applied Communication', *Applied communication in the 21st century* (1995): 79–96.
- Dencik, Lina. 'The Datafied Welfare State: A Perspective from the UK', in Hepp, Andreas, Jarke, Juliane, and Kramp, Leif (eds), *The Ambivalences of Data Power: New Perspectives in Critical Data Studies*, UK: Palgrave Macmillan, (2022):145–66.
- Ellis, Carolyn, Adams, Tony E., and Bochner, Arthur P. 'Autoethnography: An Overview', *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 12.1 (2010).
- Green, Ben, and Viljoen, Salomé. 'Algorithmic Realism: Expanding the Boundaries of Algorithmic Thought' in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, Spain: ACM, 2020: 19–31.
- Hemmings, Clare. *Why Stories Matter*, Durham: Duke University Press, 2011.
- Honig, Bonnie. *A Feminist Theory of Refusal*, Cambridge: Harvard University Press, 2021.
- Incite! Women of Color Against Violence (eds). *The Revolution Will Not Be Funded: Beyond the Non-profit Industrial Complex*, Boston: South End Press, 2007.
- Jaeggi, Rahel. *Critique of Forms of Life*, Cambridge: Harvard University Press, 2018.
- Jobin, Anna, Ienca, Marcello, and Vayena, Effy. 'The Global Landscape of AI Ethics Guidelines', *Nature Machine Intelligence* 1.9 (2019): 389–99.
- Jones, Richard G. 'Putting Privilege Into Practice Through "Intersectional Reflexivity": Ruminations, Interventions, and Possibilities', in *Reflections: Narratives of Professional Helping*, 2010.
- Lapowsky, Issie. 'How Tech Giants Court and Crush the People Who Study Them', *Protocol—The People, Power and Politics of Tech* blog, 19 March 2021. <https://www.protocol.com/nyu-facebook-researchers-scraping>.
- Metcalf, Jacob, and Moss, Emanuel. 'Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics', *Social Research: An International Quarterly* 86.2 (2019): 449–76.
- Metz, Cade and Wakabayashi, Daisuke. 'Google Researcher Says She Was Fired Over Paper Highlighting Bias in A.I.', *The New York Times*, 3 December 2020.
- Moten, Fred, and Harney, Stefan. 'The University and the Undercommons: Seven Theses', *Social Text* 22.2 (2004): 101–15.
- O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Crown, 2016.

Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge: Harvard University Press, 2016.

———. 'Machines Judging Humans: The Promise and Perils of Formalizing Evaluative Criteria.', in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2020.

Phan, Thao, et al. 'Economies of Virtue: The Circulation of 'Ethics' in Big Tech', *Science as Culture* 31.1 (2021): 1–15.

Popiel, Pawel. 'The Tech Lobby: Tracing the Contours of New Media Elite Lobbying Power', *Communication Culture & Critique* 11.4 (2018): 566–85.

Reed-Danahay, Deborah. *Auto/Ethnography: Rewriting the Self and the Social*, Oxford; New York: Routledge, 1997.

Rincón, Cami, Keyes, Os, and Cath, Corinne. 'Speaking from Experience: Trans/Non-Binary Requirements for Voice-Activated AI', in *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW1 (2021): 132:1–132:27.

Seaver, Nick. 'CARE AND SCALE: Decorrelative Ethics in Algorithmic Recommendation', *Cultural Anthropology*. 36.3 (2021): 509–37.

Shotwell, Alexis. *Against Purity: Living Ethically in Compromised Times*, Minneapolis: University of Minnesota Press, 2016.

Simpson, Audra. 'The Ruse of Consent and the Anatomy of "Refusal": Cases from Indigenous North America and Australia', *Postcolonial Studies*, 20.1 (2017): 18–33.

Sloane, Mona. 'Inequality Is the Name of the Game: Thoughts on the Emerging Field of Technology, Ethics and Social Justice', Weizenbaum Conference, Berlin, Germany, 16–17 May 2019.

Sotiropoulou, Panagiota, and Cranston, Sophie. 'Critical Friendship: An Alternative, "Care-full" Way to Play the Academic Game', *Gender, Place & Culture* (2022): 1–22.

Star, Susan Leigh. 'This is Not a Boundary Object: Reflections on the Origin of a Concept', *Science, Technology, & Human Values*, 35.5 (2010): 601–17.

Su, Norman Makoto, Lazar, Amanda, and Irani, Lilly. 'Critical Affects: Tech Work Emotions Amidst the Techlash' in *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW1 (2021): 1–27.

Tarnoff, Ben. *The Making of the Tech Worker Movement*. New York: *Logic* magazine, 2019, <https://logicmag.io/the-making-of-the-tech-worker-movement/>.

Taylor, Linnet. 'Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector', *Philosophy & Technology*, (2021):1–26.

Wagner, Ben. 'Ethics As An Escape From Regulation. From "Ethics-Washing" To Ethics-Shopping?' in *Being Profiled*, Amsterdam University Press, 2018, 84–9.

Wakabayashi, Daisuke. 'Big Tech Funds a Think Tank Pushing for Fewer Rules. For Big Tech', *The New York Times*, 24 July 2020.

Whittaker, Meredith. 'The Steep Cost of Capture', *Interactions*, 28.6 (2021): 50–5.