Press Release Ethics in AI: Performative Ethics in For-Profit AI Companies

Amanda Potasznik
Computer Science
University of Massachusetts
Boston, USA
potasznik@cs.umb.edu

2025 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS) | 979-8-3315-3228-4/25/$31.00 ©2025 IEEE | DOI: 10.1109/ETHICS65148.2025.11098431

***Abstract*— As various types of Artificial Intelligence (AI) enjoy a surge in popularity and funding, researchers, activists, and laypeople have questioned the associated negative consequences that use of AI technologies can bring. Sensitive to these concerns, many companies that specialize in various forms of AI development have published policies that describe the ethical considerations and protocols they have enacted in order to protect users and minimize societal harm that may be caused or exacerbated by the use of their products. While the policies are encouraging and seemingly comprehensive, multiple problems remain. Policy content is specific enough to encourage feelings of safety and comfort, but in many cases, it is still vague and amorphous enough to allow behavior that contradicts the initial reassurance. Companies reserve the right to change their policies as they deem appropriate, which can lead to users having outdated views on company stances. And in some cases, companies can deliberately renege on their own policies with impunity. As a result, companies can exploit users' ignorance or inattention and engage in problematic, "irresponsible" AI while maintaining a surface appearance of goodwill and ethical compliance. This critical analysis of certain AI companies' policies (and news coverage of subsequent issues and changes in those policies) aims to highlight examples of such pretense. The paper is limited to analysis of policies for 7 companies chosen for their name recognition in the AI landscape and news coverage: OpenAI, Google, X AI, Meta, Anthropic, Microsoft, and Anduril. AI companies' practice of making reassuring ethical claims in their policies then backtracking later on is labeled in this paper as "Press release ethics."**

Keywords—Artificial Intelligence ethics, Press Release ethics

I. INTRODUCTION

The public is "firmly wary" [1] about various forms of AI in their lives, and business people [2], economists [3], and Silicon Valley billionaires [4] are dutifully offering reassurance about their AI technologies. Companies that specialize in various forms of AI development ("CAIDs") pledge to develop "responsible AI" [5-8], allowing "organizations to make more ethical, effective and efficient decisions by eliminating potential sources of bias" [5]. Still, assurances that fundamental threats to our livelihoods and lives are very far down the line may ring hollow to the writers, programmers, and others whose jobs were replaced by AI within months of ChatGPT's release to the public in November 2022 [9]. Indeed, it's not uncommon for companies in general to publicly say one thing and privately do another, sometimes simultaneously, as detailed in Section II.

CAIDs show signs of engaging in the same bait-and-switch behavior when it comes to ethical concerns about their products and services. Given the specific new domain for this phenomenon, I posit a new term to describe such slippery, "ethics washing" practices: Press Release Ethics (PREs). Indeed, the ethical issues acknowledged and considerations offered by CAIDs make for a good press release: a genre of communication that is historically "self-serving" [10] and allows CAIDs to focus on the positive despite undeniable harm caused by their products [11]. Press releases enable companies to "stress good news and downplay bad news" [10], which CAIDs must do on a regular basis as the ethical concerns of AI adoption mount. The companies need to be seen as considerate of ethical concerns, but could fall behind their competitors if they were to fully embrace exclusively ethical practices, and thus are disinclined to do so from a business and, potentially, existential standpoint. PREs, in general, embody performative ethics. PREs, in the context of Responsible AI policies, are characterized by two phases. In Phase 1, companies make public ethical reassurances regarding AI applications (to appease the public). In phase 2, those companies quietly backtrack on those reassurances (to minimize harm to the company's development and profitability). I posit that many of CAIDs' purported policies are little more than ethics theater: public-facing moral principles and behind-the-scenes hypocrisy.

In addition to its obvious duplicity, this reassurance/backtracking combination in CAIDs brings its own additional ethical issues to bear. The high barrier to entry of the field of AI development, due to stratospheric demands for training data and computing power for the creation of AI programs, has resulted in researchers calling the domain "Big AI" [12]. With CAIDs enjoying so much power (and, as will be described, so little oversight), there is an inherent power imbalance in the resulting landscape. The failure of companies to adhere to their own initial policies, instead changing them to

suit evolving environments and profit goals or violating them outright, leaves ethical decision-making responsibility to third party groups and individual users (who aren't as well funded and whose concerns do not easily influence business strategies).

II. Irresponsible AI

In recognition of the disruptive and potentially destructive power of AI technologies, disparate groups have called for such technologies to be used "responsibly." That word emphasizes the duty that CAIDs have to exercise good judgment [13] in the development and deployment of AI technologies given their positions of power over the general public. Responsible AI, however, means different things to different CAIDs. A suitable definition of the term for this paper is "the framework and principles behind the design, development, and implementation of AI systems in a manner that benefits individuals, society, and businesses while reinforcing human centricity and societal value" [14]. The tenets of Responsible AI also vary across CAIDs, employees of which appear to be the authors of the most popular articles on the subject. Most CAIDs, however, claim adherence to a list of ethical considerations that is very similar to that of IBM's: "accountability, reliability, inclusion, fairness, transparency, privacy, sustainability, and governance" [15].

This list is highly ambiguous. Even the more detailed explanations about each term ("Inclusion aims to improve the usability and outcomes of machine learning models by ensuring a variety of racial, cultural, and experiential viewpoints are considered" [15]) leave plenty to the imagination. How, exactly, is that variety of viewpoints "considered"? What weight is given to various "viewpoints"? The syntax and vocabulary words themselves are bland and vague, reminiscent of a chatbot's careful prose [16]. The ambiguity of those tenets can certainly be attributed to a broad, layperson audience; not everyone concerned about AI can understand the technological processes needed to navigate, for example, IBM's impressive Fairness 360 toolkit which "includes a comprehensive set of fairness metrics for datasets and models, explanations for these metrics, and algorithms to mitigate bias in datasets and models" [17]. The layperson-friendly ambiguity could also, however, simply be an easy out; companies drawing from "a gallery of feel-good, Responsible AI technology remedies" (as one a Forbes Councils Member put it in a call for Responsible AI) [18] so that they have something to point to in order to placate uneasy users and regulators. Indeed, such bromides may provide just enough comfort for users to suppress their cognitive dissonance in using (and paying for) technologies that they find disconcerting [1]. Regardless of the CAIDs' intentions in framing and phrasing their Responsible AI pledges and policies, discrepancies between policy assurances and problematic real-world applications fit the pattern of PREs.

III. The Great Tech Company Backtracks

A. For companies in general

Companies in general sometimes find themselves obligated to backpedal on policies that favor consumers and users when it is realized that the policy is too unfavorable for the company. People who become customers based on alluring promises can find themselves on the wrong end of a bait and switch when a company backtracks on its previous policy and installs a new one, less favorable to consumers and more favorable to the continued growth of company profits. When the clothing company L.L.Bean changed its decades-old no-time-limit, no-receipt-necessary return policy to one that allowed returns only within one year of purchase, customers balked, and some even sued the company [19]. The Ford Motor Company put out a press release when it earned "a perfect score" on the "Human Rights Campaign Corporate Equality Index, a national benchmarking survey and report on corporate policies and practices related to LGBT equality in the workplace" [20], but withdrew its participation in the index a few years later when right-leaning critics voiced their displeasure of Ford's Diversity, Equity, and Inclusion (DEI) policies [21].

Of course, companies may justify such reversals as inevitable as they respond to evolving market trends, as, according to a technology consultancy firm, "even the longest-standing traditions and expectations can change overnight in an instant" [22]. If "policies remain static, trapped in a time warp of sorts," as companies keep their original promises, they may not be able to "mitigate liabilities and foster a secure operational landscape" [23] (those quotes are from a risk management company's blog article, which exemplifies an expedient use of chatbot prose itself). The business jargon belies the fact that companies draw in new customers by touting their attractive, but ultimately temporary, concerns for unwavering ethical practices, customer well-being, and fairness. Once customers are secured, the policy becomes subject to "liabilities," and logically changes to suit the companies' bottom line. PREs are irrefutably profitable, and that profitability has historically outweighed, and continues to outweigh, concern for users and consumers. In fact, some researchers have concluded that "[a]ny conversation concerned with the impact of AI and technology on society must also be concerned with the impacts of capitalism and the reality of economic profit over consumer impact" [24], putting AI companies squarely in the space of general profit-centric businesses.

Technology companies are not exempt from such profit-seeking, ethics back-tracking behavior. Google raised eyebrows when it removed the "Don't be evil" credo from the top of its company code of conduct several years ago [25]. Meta, formerly Facebook, denied rumors of listening in on users via their cell phone microphones for the purpose of serving them tailored ads multiple times since at least 2017 [26, 27], only for their advertisers to admit to the practice in 2024 [28][1]. Video telephony platform Zoom was charged by the FTC with

[1] One indication of the company's intention to engage in the behavior they stringently denied is perhaps found in their 2018 patent application for an "application module [that] record[s] ambient audio" [29]

misleading users "since at least 2016… Zoom [touted] that it offered 'end-to-end, 256-bit encryption' to secure users' communications, when in fact it provided a lower level of security" [30]. The company also made changes to its terms in 2023, quietly mentioning in the updated version that the company would now reserve the right to use customers' "video recordings, audio transcripts, or shared files… for lots of things, including training Zoom's machine learning and artificial intelligence applications" [31]. To summarize the issue,

> While each of the mega -companies, at least on paper, boasts a set of values unquestionable for their integrity, it is also true that most have had to face crises precisely due to the lack of some of the principles they themselves proclaim [32].

Similar to the ethically dynamic nature of company documentation, there is a personnel component to PREs. Some ethics-minded employees even find themselves having to permanently sever ties with the company due to blowback, as in the case of former Google researcher Timnit Gebru [33]. Facebook whistleblower Frances Haugen noted that Meta "chooses profits over safety" in documents provided to journalists before her departure from the company [34]. The performative aspect of Silicon Valley PREs is now on full display, but arguably technology companies including Meta have successfully navigated the blowback associated with the behind-the-scenes renunciation of their own purported concern for user well-being. Even when there is negative press coverage of such events, there doesn't seem to be much effect on the companies' continued operations or bottom lines [35], emphasizing both to the public and to the companies themselves that no change to their modus operandi will be demanded. This convenient repudiation of ethical tenets is now being repeated in CAIDs.

B. *For CAIDs in particular*

Certainly, CAIDs are not immune to the evolving market trends which, historically, have forced companies to relegate ethics to the backburner despite previous claims of adherence to Responsible AI policies. But the principles of technology ethics, such as the assertions that "health, safety and welfare of the public is primary" and that software should be approved

> only if [programmers] have a well-founded belief that it is safe, meets specifications, passes appropriate tests, and does not diminish quality of life, diminish privacy or harm the environment. The ultimate effect of the work should be to the public good" [36],

are categorized as "fundamental" and certainly should not change with market trends. Yet CAIDs arguably not only break from moral codes in their business practices, but also break their own policies as the winds shift. Investigative reporting from the New York Times showed that "OpenAI, Google and Meta ignored corporate policies, altered their own rules and discussed skirting copyright law as they sought online information to train their newest artificial intelligence systems" [37]. It seems that CAIDs are ready to adhere to publicly declared Responsible AI tenets until it's no longer profitable to do so - a clear match to the description of PREs.

Around the end of 2022, multiple CAIDs united in a public call for a moratorium on AI development [38], but the pause never quite materialized [39]: another indication of the variance between CAIDs stated plans and their own actions[2]. Individual CAID safety teams also seem to suffer startling rates of attrition, and in cases total disbandment, which will be discussed in detail in the next section.

1) Open AI: The formation of "safety teams" at OpenAI was heralded with press coverage and touted on the company website [43]. The original team was disbanded, however, after less than one year in existence [44], [45] and replaced by one headed by its own CEO [46][3]. Arguably, the precariousness of safety teams at OpenAI is due in part to "a culture of recklessness and secrecy" as one former employee described [48]. Journalists have struggled to find former OpenAI employees who were willing to share their experiences due to some punitive protocols for those leaving the company:

> That's partly because OpenAI is known for getting its workers to sign offboarding agreements with non-disparagement provisions upon leaving. If you refuse to sign one, you give up your equity in the company, which means you potentially lose out on millions of dollars. [45]

Still, with some current employees speaking anonymously and some former employees forgoing the offboarding agreement, there are records of people with knowledge of the company's inner workings reporting concerns that do not align with the upbeat assurances we see in the OpenAI safety policy. Daniel Kokotajlo is "a former researcher in OpenAI's governance division" who said in an interview "that the probability that advanced A.I. will destroy or catastrophically harm humanity… is 70 percent" [48]. The alignment of OpenAI's safety practices with the characterization of PREs is also evident in consideration of the former researcher's description of a disconnect between company safety protocols and actions: "At OpenAI, Mr. Kokotajlo saw that even though the company had safety protocols in place … they rarely seemed to slow anything down" [48].

Beyond the uses of such technologies, their creation is also facilitated by duplicitous practices, including scraping training data from unapproved sources. OpenAI created speech to text programs so that they could feed the text from YouTube video transcriptions into their LLMs. Not even Google, though, could bring themselves to call OpenAI out on this practice. Why not?

---

[2] Relying on CAIDs to regulate themselves is unsurprisingly problematic, but their reaction to government regulation is also difficult to follow. Some CAIDs have publicly supported state-level legislation in California that puts limits on their ability to train their models, while others have opposed it [40]. The legislation has been criticized both for not going far enough in its limitations of CAIDs and for going too far and "stifling innovation" according to OpenAI [41]. There is no federal level legislation regulating CAIDs at the time of this writing [42].

[3] Exemplifying the dizzying pace of change not only of CAID technology but also company structure, OpenAI CEO Sam Altman stepped down from that position a few weeks later, "after five U.S. senators raised questions about OpenAI's policies in a letter addressed to Altman this summer." [47]

> Some Google employees were aware that OpenAI had harvested YouTube videos for data, two people with knowledge of the companies said. But they didn't stop OpenAI because Google had also used transcripts of YouTube videos to train its A.I. models, the people said. That practice may have violated the copyrights of YouTube creators. So if Google made a fuss about OpenAI, there might be a public outcry against its own methods, the people said. [37]

CAIDs, then, cannot call out unethical behavior or uses of their datasets by other companies because they themselves may be engaged in the same secretive and duplicitous practices. As such, phase 2 of PREs (that is, the phase in which the company backpedals on its previous declarations of ethical behavior) are understandably played down. Only investigative journalism and research can bring the actual business practices to light, since the CAIDs themselves maintain the charade of adhering to their own policies.[4]

2) Google (Gemini): In an open letter introducing Google's AI assistant, Gemini, Google CEO Sundar Pichai wrote, "We're approaching this work boldly and responsibly" [49] and alluded the company's AI principles, which state that Google's objectives for AI applications will "[b]e socially beneficial, be built and tested for safety, be accountable to people" and that "we will not design or deploy AI [that is applied in the area of t]echnologies that cause or are likely to cause overall harm [6]. The company goes beyond what most of its AI competitors provide in the way of documentation of changes to policies, providing a list of updates and iterations of its policies in easily accessible PDF form. When initial versions of Gemini were documented to generate images with historical inaccuracies and racial bias [50], the program was suspended and the senior vice president of the company published an open letter entitled "Gemini image generation got it wrong. We'll do better" [51]. The overall impression when reading the company's Gemini introduction and AI Principles is one of integrity: taking responsibility for its formidable computing power and full disclosure of its shortfalls.

But beyond the company's own documentation, PRE attributes are evident. For example, it was found that "Google violated its promised standards" in placing third-party advertisements [53], and since the company uses AI for advertising[5], this violation fits comfortably in the AI PRE category: Google promised one thing in its AI policies but did another, accepting payments from customers who thought the company was following its own rules. Summarizing a logical response to such PREs, one customer said "'I feel cheated… What I requested to buy was not what I got'" [53].

The size needed for viable (and marketable) LLM training sets has also led CAIDs, including Google, to reconsider its previous guarantees of privacy and protection for users. Those size requirements have already resulted in CAIDs exhausting publicly available data online and bumping up against paywalled content in their search for more data to feed their programs [55]. I have already discussed OpenAI and Google scraping user content from the open internet in the form of YouTube videos. Google has an advantage in the realm of information beyond the open internet, of course. Millions of people use its Drive services to create their own documents, and its review platform also contains plenty of user-generated content. The company has now changed its terms of service (after the vast majority of users had agreed to previous versions with different clauses) to enable it to harvest those sources of personal, private information for the purpose of training its LLMs:

> Last year, Google also broadened its terms of service. One motivation for the change, according to members of the company's privacy team and an internal message … was to allow Google to be able to tap publicly available Google Docs, restaurant reviews on Google Maps and other online material for more of its A.I. products. [37]

"Tapping" information sources beyond the scope of their initial protocols is a glaring example of secondary use, or the use of data for reasons beyond their primary purpose. In this case, Google markets (and profits from) its Drive services as allowing users to easily generate documents for themselves, an arrangement that has drawn customers since 2012 [56]. Only recently has the company maneuvered so that it can now also harvest its customers' documents to train AI models, constituting secondary use, the exclusive benefit of which is derived by the company.

In chorus with other CAIDs, Google has also suffered attrition and full disbandment of their AI safety team, the "Responsible Innovation team" [58]. Also in chorus with many CAIDs, the company has publicly declared that despite the safety team disbanding, safety is a top priority for the company: "Despite these changes, a Google spokesperson assured that the team's mission will proceed in an even more robust manner, though specifics were not disclosed" [57]. It is the lack of specifics, of course, that are emblematic of PREs: the overarching public message is one of reassurance, but details don't bear that out.

Perhaps most egregiously, Google quietly changed course on its pledge to develop AI for responsible uses, with a gentle notice at the top of its 2018 blog post about responsible AI: "We've made updates to our AI Principles." Those updates?

> The company removed language promising not to pursue "technologies that cause or are likely to cause overall harm," "weapons or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people," "technologies that gather or use information for surveillance violating internationally accepted norms," and "technologies whose purpose contravenes widely accepted principles of international law and human rights." [59]

The adherence to PRE methods is undeniably apparent in this

[4] Since the writing of this paper, OpenAI indulged in more PREs when, in December 2024, they partnered with Anduril Industries [103]. That company is analyzed in part 7 of this list.

[5] "Google uses AI to set the right bids, reach the right searches, and create the most relevant ad for consumers" [54]

removal.

3) Grok-2 (via xAI:) The social media company X, run by Elon Musk, has also entered the generative AI space with Grok (and now Grok-2). Musk has a rather negative record when it comes to business ethics: "Elon Musk has violated a lot of the social contract, and its basis of trust, with employees, with investors, suppliers, regulators, and other parts of his ecosystem" [60]. Specifically, the incentivization of violent content on X under Musk has drawn extensive criticism [60], and he has been charged with fraud by the Securities and Exchange commission [62]. The CEO doesn't sugarcoat his views on the damaging potential of AI: "There is some chance that is above zero that AI will kill us all… I think we should also consider the fragility of human civilization" [63]. Given this fairly negative public perception, it is tempting to spare Grok-2 from the list of PRE CAIDs. But despite a CEO with plenty of bad press, there are still vestiges of performative ethical assurances scattered throughout the company's documentation.

In the Grok-2 "About" page, for example, it is claimed that the CAIDs goals are for the common good: "xAI is a company working on building artificial intelligence to accelerate human scientific discovery. We are guided by our mission to advance our collective understanding of the universe." The Acceptable Use Policy for Grok-2, composed of only 248 words including the title and date, concludes with a phrase that can only be interpreted as its writers claiming the high moral ground: "Be a good human, it's really not that hard" [64].

As such, there is indeed the hallmark discrepancy between stated company policy and documented company action. "Unlike more restrained AI models like ChatGPT or Google's Gemini, Grok-2 seems to operate with fewer ethical guardrails" [65]. Grok-2, born as it is from a social media company, joins Meta as a CAID that has direct influence in global discourse. Rather than using that influence to attempt to slow the waves of online misinformation that plague society today, AI models like Grok-2 instead add to the chaos: "In an era when distinguishing fact from fiction online is already challenging, tools like Grok-2 could exacerbate the spread of misinformation and deepen societal divisions" [65]. Of course, X and Grok-2 are not the only social media sites that profit from users' engagement with manufactured content and the calcification of viewpoints in algorithm-powered filter bubbles. Other social media companies, however, at least go through the motions of content moderation. "Grok allows misinformation to proliferate unchecked, a significant departure from the moderated environments that its competitors maintain" [66].

4) Meta AI: As one of the most successful social networking platforms in existence, Meta (a parent company that now owns Facebook, Instagram, and Whatsapp) has substantial name recognition when it comes to dubious ethical practices. Research documenting problematic business practices at the company goes back decades [67] and continues to this day.

Meta's ethical messaging is, predictably, heavy on optimistic rhetoric and world improvement, from its introduction ("we build products and experiences to give people the power to build community and bring the world closer together") to the individual principles ("Keep people safe and protect privacy—we are committed to protecting our communities from harm") [68]. But true to PRE form, these policies are more aspirational than concrete. Meta founder Mark Zuckerberg has testified at governmental hearings that were called in response to legislators' alarm at the company's lucrative algorithms and the myriad societal ills that came as a direct result: the proliferation of misinformation, societal division, and even child exploitation, among others [69]. It is widely acknowledged that the company's ethical failures are, to borrow from programmer Sandra Lee Harris' 1971 user manual [70], a feature, not a bug: "If Facebook employed a business model focused on efficiently providing accurate information and diverse views, rather than addicting users to highly engaging content within an echo chamber, the algorithmic outcomes would be very different" [71].

As Meta joins the CAID space, the company is further engaging in PRE practices. The discrepancy between company policy and business practice are clear. "Privacy-focused" messaging platform WhatsApp shares users' data with Facebook and gradually degrades if users do not agree to that policy [72]. Zuckerberg is also laying the foundation to abdicate ethical responsibility from his own company to the U.S. government when it comes to controlling the AI technologies from which he profits: "Congress should engage with AI to support innovation and safeguards. This is an emerging technology, there are important equities to balance here, and the government is ultimately responsible for that" [63].

While it seems that Zuckerberg believes Meta's limitations should come from Congress, not its own policies, company lawyers reportedly do not shy away from lawsuits that may result from sidestepping government regulations. In its quest to harvest training data for its AI programs, the company appears ready to buy out obstacles in its path and to break some rules in the name of expediency:

> At Meta, which owns Facebook and Instagram, managers, lawyers and engineers last year discussed buying the publishing house Simon & Schuster to procure long works, according to recordings of internal meetings obtained by The Times. They also conferred on gathering copyrighted data from across the internet, even if that meant facing lawsuits. Negotiating licenses with publishers, artists, musicians and the news industry would take too long, they said. [37]

Given the discrepancies between Meta's own policies and their past and current behavior, it is reasonable to conclude that they will fit squarely into the PRE category among CAIDs.[6]

[6] Since the writing of this paper, Meta and its AI model Llama have made headlines for backtracking on their previous policy that excluded military use from its terms. The company has now made the models available for military use only for the U.S., Canada, Great Britain, New Zealand, and Australia. In doing so they made Llama open source, a move that quickly resulted in its use by the Chinese military [101], [102].

5) Anthropic (Claude): The company that developed the Claude AI assistant, Anthropic, has cultivated a reputation of ethical behavior amidst rival AI giants [73]. Its founders all started at OpenAI, but left to create their own AI company "on a promise of building reliable and steerable AI systems" [74]. When more employees and executives left OpenAI due to ethical concerns (documented in the previous section), several joined Anthropic [75], with the implied message that Anthropic did not share the dismissive attitude toward ethics and safety that OpenAI had. Indeed, Anthropic researchers have identified and publicized shortcomings of their LLMs [76], and the company co-president has been profiled in The Atlantic in an article entitled "Building AI With a Conscience" [77]. The company declares on their help center website that "User safety is core to Anthropic's mission of creating reliable, interpretable, and steerable AI systems" [78], and their AI Constitution details an approach to training models while "helping to avoid toxic or discriminatory outputs, avoiding helping a human engage in illegal or unethical activities, and broadly creating an AI system that is helpful, honest, and harmless" [79].

But a closer look reveals some hallmarks of PREs even in this safety-focused company. Some subtle gray-on-white text at the top of *The Atlantic* article indicates that it is "Sponsor Content" rather than a feature created by the editorial staff[7]. Anthropic accepted a multi-billion-dollar investment from Google, resulting in an antitrust investigation in the United Kingdom [80]. Web publishers have bristled at the company's training methods, relating that "Anthropic is swarming their sites and ignoring their instructions to stop collecting their content to train its model" [81]. The contradictions between this CAID's stated practices and its real-world practices have led journalists to ask: "Is it even possible to run an AI company that advances the state of the art while also truly prioritizing ethics and safety?" and ultimately answer in the negative, concluding that "even high-minded Anthropic is becoming an object lesson in that impossibility" [82].

6) Microsoft (Copilot): Like most CAIDs, Microsoft extensively documents its commitment to ethical business practices on its consumer-facing site:

> Our commitment to corporate responsibility and integrity guides everything we do as a company and defines the work of our ethics and compliance program. We have high ethical standards governing the way we conduct our business, standards that we also apply to our suppliers and business partners. Our business practices and standards reflect our commitment to making a positive impact around the globe. We demand such high standards from ourselves and our partners to preserve trust with our customers, governments, investors, partners, representatives, and each other, and because it is the right thing to do.

Additionally, on the company's "Trust Code" documentation for employees (linked to from the user-facing page), a large, bold font subheading instructs: "When making decisions, ask yourself: does this build or harm trust with our customers?" There are 53 pages in the Trust Code document, detailing the company's supposed commitment to earning trust with customers, governments, communities, investors, the public, suppliers, and employees. All told, there are dozens of pages of company documentation dedicated to the message that Microsoft is an ethical, trustworthy company. As another researcher summarizes: "Overall, Microsoft's AI principles reflect a comprehensive and thorough approach to the development and use of their own technology" [32].

As with other CAIDs, some of those claims appear to be aspirational rather than de facto. Researchers have called the company out for unethical business practices for decades, documenting issues from unfair competitive practices [82] to enabling censorship by complying with Chinese requirements for doing business [84]. In the AI realm, in 2023, the company

> laid off its entire ethics and society team within the artificial intelligence organization… [leaving] Microsoft without a dedicated team to ensure its AI principles are closely tied to product design at a time when the company is leading the charge to make AI tools available to the mainstream" [85]

A software engineer at the company reported that Microsoft's AI image generator "created … sexualized images of women in violent tableaus, and underage drinking and drug use" [86]. While it arguably takes some trial and error to discover and fix such issues, the problem for Microsoft appears to be at the structural level. Firing the ethics and society team, argue some ethicists, resulted in issues, which should have been fixed prior to release, being released into the world in real time: "Microsoft is in deep trouble because of the model they've adopted. And the ethicists who are pulling the whistleblower siren have to do so because they got rid of the people inside who would help them" [87].

7) Anduril: Anduril, a company that makes autonomous weapons (uniting AI and warfare), meets the qualifications for PREs. A journalist interviewing one of the company's co founders and former Trump military advisor, Trae Stephens, summed up the ethical bait-and-switch in his questioning: "When I wrote about Anduril in 2018, the company explicitly said it wouldn't build lethal weapons. Now you are building fighter planes, underwater drones, and other deadly weapons of war. Why did you make that pivot?" [87]. The cofounder defended his company's ethical flip-flop with vague business jargon and xenophobic fear-mongering, and continued to claim that the company is standing by their ethical principles in creating machines that kill human beings:

> We responded to what we saw, not only inside our military but also across the world. We want to be aligned with delivering the best capabilities in the most ethical way possible. The alternative is that someone's going to do that anyway, and we believe that we can do that best. [87]

[7] "This content is made possible by our sponsor and is independent of The Atlantic's editorial staff" [76]

"We can do that best" is a slogan worthy of a press release. By invoking that phrase as he gives reasons why, six years ago, he said the company would do the opposite of what it's doing today, Stephens gave a prime example of PREs.

Stephens also echoed Zuckerberg's previously documented abdication of ethical responsibility, concluding that the government should set ethical boundaries rather than companies themselves: "I don't think that there's a whole lot of utility in trying to set our own line when the government is actually setting that line" [87]. His fellow co-founder Palmer Luckey agreed, stating "I don't think I'm the guy to teach people ethics. I can give people my perspective" [88][8].

Later in the interview, Stephens pushed the limits of performative ethics, declaring that not only is his autonomous weapons company doing the right thing, but that he wishes other technology companies would be more ethical, going so far as to claim the ultimate divine ethical endorsement: "The call that I have been trying to make to the tech community is that we have a moral obligation to do things to benefit humanity, to draw us closer to God's plan for his people" [89].

Anduril does not have a published ethics code to contradict. I posit, however, that the company still qualifies as engaging in PREs due to its co-founders going back on their word about developing autonomous weapons in the first place, their hypocrisy in suggesting that other technology companies have scorned Anduril's pleas for better moral behavior, and their blasphemous audacity in declaring that they are enacting God's plan by profiting from the creation of autonomous killing machines.

IV. Performative Ethics in Academia and Research

Academic research has also been accused of such performative ethics and cannot be excluded from this conversation. While researchers may consider themselves to be neutral parties in the documentation and expansion of AI ethics, their lack of introspection and self-analysis is performative and potentially destructive: "Without systemic analysis… work dedicated to positively improving the impact of technology on society will be performative at best and reify systems of oppression at worst" [24]. Critical theory scholars have identified systems of oppression in myriad places in society, and now those systems are manifesting in a post-AI world.

It is easy to see how systems of oppression may manifest in AI companies: CAIDs oppress regular users by training systems on their content without notice or consent [12], [90], and have used the resulting technology to render users' jobs obsolete [91, 92] and even kill people [103]. While CAIDs profit and grow from that oppression, regular users are encouraged to adapt to the situation [93]. The capitalistic component of our society seems to have convinced regular people that they must use (i.e., pay subscription fees to use) AI in order to remain employable in a job market in which that same AI threatens their employability. This paradox, which greatly favors CAIDs at the expense of regular people, is indicative of the creeping normalcy phenomenon:

> … the term "creeping normalcy" [can] refer to such slow trends concealed within noisy fluctuations. If the economy, schools, traffic congestion, or anything else is deteriorating only slowly, it's difficult to recognize that each successive year is on the average slightly worse than the year before, so one's baseline standard for what constitutes "normalcy" shifts gradually and imperceptibly. It may take a few decades of a long sequence of such slight year-to-year changes before people realize, with a jolt, that conditions used to be much better several decades ago, and that what is accepted as normalcy has crept downwards. [94]

Indeed, even in the relatively quick-paced explosion of the use of certain types of AI, CAIDs have established the normalcy, and inevitability, of their domination in society.

As mentioned in the introduction, due to the vast amounts of training data and computing power needed to develop AI, researchers have classified CAIDs as "Big AI." Smaller and slower moving companies (and perhaps those that pause development in order to ensure ethical goals are met) are simply shut out. This setup contextualizes the CAID sentiment that AI that is developed *fast* is the most profitable, an idea summarized by Michael Woolridge of Oxford. In an article he wrote about Big AI, Woolridge theorized about the inner monologue of a CAID leader, writing "the race for scale that we have witnessed in AI over the past five years is perhaps no surprise: if bigger is better, then let us make it bigger—and let us do it before our competitors" [12]. Hence the importance of groups of resistance: without researchers, advocacy groups, and the government reining in CAIDs, only capitalism and its mantra of constant growth will control CAIDs, which of course means no control at all. Their dominance will continue to breed more success, contributing to a status quo in which ethical concerns are effectively performative: lines in a script that do nothing to control or mitigate profitable harm on a large scale. The status quo is a powerful force in upholding systems of oppression [95], within tech fields and without.

Researchers have documented user trust in big companies, showing that many users use websites and applications without reading the terms and conditions because they believe that "If there were anything bad in the policy, it would be illegal" [96]. But governmental regulation is far outpaced by CAID developments, and there is still no enforceable law at the federal level that regulates CAIDs. Instead, there is a patchwork set of state laws (in Oregon, Montana, New Hampshire, Tennessee, and Delaware) [97] and national blueprints [98] that indicate an understanding of the issue but not a concrete way to address it or punish violations at the federal level.

With CAIDs being profit-motivated, consumers feeling the need to adapt to CAIDs, and the U.S. government lagging in enforceable legislation to temper unethical AI practices, it is arguably researchers who are best placed to identify and

[8] That perspective has historically been objectively harmful, as evidenced by Luckey "secretly funding Nimble America, a 501(c)4 'social welfare' organization responsible for generating white supremacist memes and lobbying for a Trump presidency" [88]

interrogate problematic AI ethics policies and actions. I add another consideration to McFadden and Alvarez's call for researcher critical theory: research that addresses ethics in AI without interrogating the adjacent PREs component contributes to the tacit acceptance, and subsequent perpetuation, of historical systems of oppression. Research of CAID PREs without denouncement adds to the creeping normalcy mentioned previously, and contributes to inertia bias: When the first wave of users accepts the discrepancies between CAIDs stated policies and their actual practices, the acceptance becomes entrenched and harder to resist for future generations.

V. Conclusion: Optimism and Healthy Skepticism

Subjecting new AI technologies to critical analysis does not stop their exponential rate of development and adoption in the marketplace. As described in section II, even the CAIDs themselves find controlling or decelerating AI development an impossible task. This paper focused on seven CAIDs: OpenAI, Google, X AI, Meta, Anthropic, Microsoft, and Anduril. That focus should not, however, be interpreted to mean that other CAIDs don't exist or don't have problematic PREs in their policies.

Evolving landscapes are, of course, inevitable as people use brand-new and untested technologies, including those that use AI. People and companies may be aspirational and have to regroup when they realize their stated goals are too optimistic and not 100% compatible with the current reality. The fact that CAIDs and other businesses fall short in similar ways may be part of the human experience. The medical breakthroughs alone of AI use demonstrate that a complete shutdown of this new technology would be damaging to humans.

But given the power, scale, and amounts of money flowing into the sector, now is the time to be brutally honest about what CAIDs do. There are glaring discrepancies between CAID companies' policies and their actual practices, ranging in severity from copyright violations to killing human beings. Researchers must amplify their concerns about the ethical challenges of these systems and companies. Teachers must resist calls to implement AI in their classrooms without a healthy dose of interrogation, and perhaps even reject those calls altogether in order to ensure that students do not blindly adopt such tools for the everyday challenges they face. Governments should regulate CAIDs heavily. CAIDs themselves must abandon PREs and instead make their business practices match their policies (and vice versa).

With governments slow to develop legislative limits to CAID power, and CAIDs themselves insisting that their only limits should come from governmental regulation, researchers and users may be the best advocates for change. As matters stand, the most urgent primary need for researchers and users is a strong sense of cynicism when reading ethics statements and policies from CAIDs. Given the volatile relationship between profit-driven business practices and ethical concerns, we cannot take CAIDs at their word until there is a dramatic shift in accountability for the industry.

References

[1] J. Ray, "Americans Express Real Concerns About Artificial Intelligence," Gallup.com. Accessed: Sep. 19, 2024. [Online]. Available: https://news.gallup.com/poll/648953/americans-express-real-concerns-artificial-intelligence.aspx

[2] E. Council Forbes Technology, "Council Post: Why You Don't Need To Worry About AI," Forbes. Accessed: Sep. 19, 2024. [Online]. Available: https://www.forbes.com/sites/forbestechcouncil/2017/03/15/why-you-dont-need-to-worry-about-ai/

[3] The Economic Times Bureau, "Don't worry (too much) about AI," The Economic Times, Jun. 06, 2023. Accessed: Sep. 19, 2024. [Online]. Available: https://economictimes.indiatimes.com/opinion/et-editorial/dont-worry-too-much-about-ai/articleshow/100801749.cms?from=mdr

[4] K. Leswing, "Bill Gates explains why we shouldn't be afraid of A.I.," CNBC. Accessed: Sep. 19, 2024. [Online]. Available: https://www.cnbc.com/2023/07/12/bill-gates-explains-why-we-shouldnt-be-afraid-of-ai.html

[5] Accenture, "Responsible AI Governance Consulting & Solutions | Accenture," Accenture. Accessed: Sep. 10, 2024. [Online]. Available: https://www.accenture.com/us-en/services/data-ai/responsible-ai

[6] Google, "Google Responsible AI Practices," Google AI. Accessed: Sep. 10, 2024. [Online]. Available: https://ai.google/responsibility/responsible-ai-practices/

[7] IBM, "What is responsible AI? | IBM," IBM. Accessed: Sep. 10, 2024. [Online]. Available: https://www.ibm.com/topics/responsible-ai

[8] Microsoft, "Microsoft Responsible AI Standard General Requirements." Accessed: Sep. 26, 2024. [Online]. Available: https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf

[9] OpenAI.com, "Introducing ChatGPT," OpenAI. Accessed: Sep. 10, 2024. [Online]. Available: https://openai.com/index/chatgpt/

[10] E. Guillamon-Saorin, B. G. Osma, and M. J. Jones, "Opportunistic disclosure in press release headlines," Account. Bus. Res., Jun. 2012, Accessed: Sep. 26, 2024. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00014788.2012.632575

[11] D. Acemoglu, "Harms of AI," Sep. 2021, National Bureau of Economic Research: 29247. doi: 10.3386/w29247.

[12] M. Wooldridge, "Welcome to Big AI," IEEE Intell. Syst., vol. 37, no. 3, pp. 24–26, May 2022, doi: 10.1109/MIS.2022.3184429.

[13] Cambridge Dictionary, "Responsibly." Accessed: Sep. 26, 2024. [Online]. Available: https://dictionary.cambridge.org/us/dictionary/english/responsibly

[14] A. Lawson, "AI vs. Responsible AI: Why is it Important?," Responsible AI. Accessed: Sep. 26, 2024. [Online]. Available: https://www.responsible.ai/ai-vs-responsible-ai-why-is-it-important/

[15] E. Sanders, "Understanding Responsible AI," Neudesic. Accessed: Sep. 26, 2024. [Online]. Available: https://www.neudesic.com/blog/understand-responsible-ai/

[16] J. Brady, M. Kuvalja, A. Rodrigues, and S. Hughes, "Does ChatGPT make the grade?," Feb. 2024, doi: 10.17863/CAM.106034.

[17] R. K. E. Bellamy et al., "AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias," arXiv.org. Accessed: Sep. 26, 2024. [Online]. Available: https://arxiv.org/abs/1810.01943v1

[18] T. Robinson, "Imagine Irresponsible AI To Help Define Your Responsible AI Strategy," Forbes. Accessed: Sep. 26, 2024. [Online]. Available: https://www.forbes.com/councils/forbestechcouncil/2023/08/07/imagine-irresponsible-ai-to-help-define-your-responsible-ai-strategy/

[19] J. Abbey, M. Ketzenberg, and R. Metters, "A More Profitable Approach to Product Returns," MIT Sloan Manag. Rev., vol. 60, no. 1, pp. 1–6, Fall 2018.

[20] "Ford Received Perfect Score on Human Rights Campaign 2017 Corporate Equality Index | Ford Media Center." Accessed: Oct. 03, 2024. [Online]. Available: https://media.ford.com/content/fordmedia/fna/us/en/news/2016/12/05/ford-perfect-score-human-rights-index.html

[21] E. Yildirim, "Ford joins list of companies walking back DEI policies," CNBC. Accessed: Oct. 03, 2024. [Online]. Available: https://www.cnbc.com/2024/08/28/ford-joins-list-of-companies-walking-back-dei-policies.html

[22] P. Ramesh, "Adapt or Fail: Evolving Leadership in a Hybrid World," Peterson Technology Partners. Accessed: Oct. 03, 2024. [Online]. Available: https://www.ptechpartners.com/2023/01/31/adapt-or-fail-evolving-leadership-in-a-hybrid-world/

[23] J. DiChiara, "Company Policies: The Need for Regular Review and Revision," SAI360. Accessed: Oct. 03, 2024. [Online]. Available:

https://www.sai360.com/resources/grc/company-policies-the-need-for-regular-review-and-revision-blog

[24] Z. McFadden and L. Alvarez, "Performative Ethics From Within the Ivory Tower: How CS Practitioners Uphold Systems of Oppression," J. Artif. Intell. Res., vol. 79, pp. 777–799, Mar. 2024, doi: 10.1613/jair.1.15423.

[25] K. Conger, "Google Removes 'Don't Be Evil' Clause From Its Code of Conduct," Gizmodo. Accessed: Oct. 03, 2024. [Online]. Available: https://gizmodo.com/google-removes-nearly-all-mentions-of-dont-be-evil-from-1826153393

[26] A. Hern, "No, Facebook isn't spying on you. At least not with the microphone," The Guardian, Nov. 09, 2017. Accessed: Oct. 03, 2024. [Online]. Available: https://www.theguardian.com/technology/2017/nov/09/facebook-spying-on-you-microphone-creepy-data-conspiracy-theories

[27] A. Hern, "Facebook denies eavesdropping on conversations to target ads, again," The Guardian, Oct. 30, 2017. Accessed: Oct. 03, 2024. [Online]. Available: https://www.theguardian.com/technology/2017/oct/30/facebook-denies-eavesdropping-on-conversations-to-target-ads-again

[28] N. Al-Sibai, "In Leak, Facebook Partner Brags About Listening to Your Phone's Microphone to Serve Ads for Stuff You Mention," Futurism. Accessed: Oct. 03, 2024. [Online]. Available: https://futurism.com/the-byte/facebook-partner-phones-listening-microphone

[29] A. M. Husain and Y. Xu, "Broadcast Content View Analysis Based on Ambient Audio Recording," 20180167677, Jun. 14, 2018 Accessed: Oct. 03, 2024. [Online]. Available: https://www.freepatentsonline.com/20180167677.html

[30] J. Brodkin, "Zoom lied to users about end-to-end encryption for years, FTC says," Ars Technica. Accessed: Oct. 10, 2024. [Online]. Available: https://arstechnica.com/tech-policy/2020/11/zoom-lied-to-users-about-end-to-end-encryption-for-years-ftc-says/

[31] D. P. Williams, "Zoom Became a Part of Daily Life. It Needs to Tell Users Exactly How It's Using Their Data," Wired, Aug. 10, 2023. Accessed: Oct. 10, 2024. [Online]. Available: https://www.wired.com/story/zoom-became-a-part-of-daily-life-it-needs-to-tell-users-exactly-how-its-using-their-data/

[32] F. Morandín-Ahuerma, "Ethics of AI from global companies: Microsoft, Google, Meta, and Apple," 2023, pp. 137–161.

[33] K. Hao, "We read the paper that forced Timnit Gebru out of Google. Here's what it says.," MIT Technology Review. Accessed: Oct. 03, 2024. [Online]. Available: https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/

[34] R. Mac and C. Kang, "Whistle-Blower Says Facebook 'Chooses Profits Over Safety' - The New York Times." Accessed: Oct. 03, 2024. [Online]. Available: https://www.nytimes.com/2021/10/03/technology/whistle-blower-facebook-frances-haugen.html

[35] S. Dixon, "Meta: annual revenue and net income 2023," Statista. Accessed: Oct. 03, 2024. [Online]. Available: https://www.statista.com/statistics/277229/facebooks-annual-revenue-and-net-income/

[36] D. Gotterbarn, K. Miller, and S. Rogerson, "Software engineering code of ethics," Commun ACM, vol. 40, no. 11, pp. 110–118, Nov. 1997, doi: 10.1145/265684.265699.

[37] C. Metz, C. Kang, S. Frenkel, S. A. Thompson, and N. Grant, "How Tech Giants Cut Corners to Harvest Data for A.I.," The New York Times, Apr. 06, 2024. Accessed: Oct. 08, 2024. [Online]. Available: https://www.nytimes.com/2024/04/06/technology/tech-giants-harvest-data-artificial-intelligence.html

[38] R. Heath, "It's six months after that open letter calling for a 'six-month pause' in AI work," Axios. Accessed: Sep. 10, 2024. [Online]. Available: https://www.axios.com/2023/09/22/ai-letter-six-month-pause

[39] W. Knight, "Six Months Ago Elon Musk Called for a Pause on AI. Instead Development Sped Up | WIRED." Accessed: Sep. 10, 2024. [Online]. Available: https://www.wired.com/story/fast-forward-elon-musk-letter-pause-ai-development/

[40] I. F. Gold Ashley, "California's AI safety bill is dividing big tech," Axios. Accessed: Oct. 16, 2024. [Online]. Available: https://www.axios.com/2024/08/28/california-ai-regulation-bill-divides-tech-world

[41] M. Morrone, "Exclusive: Current and former employees of leading AI firms support California's AI bill," Axios. Accessed: Oct. 16, 2024. [Online]. Available: https://www.axios.com/2024/09/09/openai-anthropic-deepmind-employees-ai-bill

[42] M. Curi and A. Gold, "AI regulation hits a wall in the House," Axios. Accessed: Oct. 16, 2024. [Online]. Available: https://www.axios.com/pro/tech-policy/2024/06/18/ai-regulation-hits-a-wall-in-the-house

[43] "Introducing Superalignment." Accessed: Oct. 03, 2024. [Online]. Available: https://openai.com/index/introducing-superalignment/

[44] W. Knight, "OpenAI's Long-Term AI Risk Team Has Disbanded," Wired, May 17, 2024. Accessed: Oct. 03, 2024. [Online]. Available: https://www.wired.com/story/openai-superalignment-team-disbanded/

[45] S. Samuel, "'I lost trust': Why the OpenAI team in charge of safeguarding humanity imploded," Vox. Accessed: Sep. 10, 2024. [Online]. Available: https://www.vox.com/future-perfect/2024/5/17/24158403/openai-resignations-ai-safety-ilya-sutskever-jan-leike-artificial-intelligence

[46] E. Roth, "OpenAI has a new safety team — it's run by Sam Altman," The Verge. Accessed: Oct. 03, 2024. [Online]. Available: https://www.theverge.com/2024/5/28/24166105/openai-safety-team-sam-altman

[47] K. Wiggers, "Sam Altman departs OpenAI's safety committee," TechCrunch. Accessed: Oct. 03, 2024. [Online]. Available: https://techcrunch.com/2024/09/16/sam-altman-departs-openais-safety-committee/

[48] K. Roose, "OpenAI Insiders Warn of a 'Reckless' Race for Dominance," The New York Times, Jun. 04, 2024. Accessed: Oct. 03, 2024. [Online]. Available: https://www.nytimes.com/2024/06/04/technology/openai-culture-whistleblowers.html

[49] S. Pichai and D. Hassabis, "Introducing Gemini: our largest and most capable AI model," Google. Accessed: Oct. 04, 2024. [Online]. Available: https://blog.google/technology/ai/google-gemini-ai/

[50] T. Spangler, "Google Suspends AI Tool's Image Generation of People After It Created Historical 'Inaccuracies,' Including Racially Diverse WWII-Era Nazi Soldiers," Variety. Accessed: Oct. 04, 2024. [Online]. Available: https://variety.com/2024/digital/news/google-gemini-ai-image-racial-inaccuracies-nazi-soldiers-1235919168/

[51] P. Raghavan, "Gemini image generation got it wrong. We'll do better.," Google. Accessed: Oct. 04, 2024. [Online]. Available: https://blog.google/products/gemini/gemini-image-generation-issue/

[52] H. Simons, "Has Google done anything unethical? Gemini changes its answer mid-sentence," Android Authority. Accessed: Oct. 04, 2024. [Online]. Available: https://www.androidauthority.com/google-gemini-unethical-change-answer-3453913/

[53] P. Haggin, "Google Violated Its Standards in Ad Deals, Research Finds," Wall Street Journal, Jun. 27, 2023. Accessed: Oct. 04, 2024. [Online]. Available: https://www.wsj.com/articles/google-violated-its-standards-in-ad-deals-research-finds-3e24e041

[54] "Your guide to AI-powered Search ads - Google Ads Help." Accessed: Oct. 04, 2024. [Online]. Available: https://support.google.com/google-ads/answer/12158267?hl=en

[55] S. Longpre et al., "Consent in Crisis: The Rapid Decline of the AI Data Commons," 2024.

[56] K. Finley, "Google Apps Now Offers Business Process Automation on Google Sites with Scripts," ReadWriteWeb, Oct. 2010, [Online]. Available: http://www.readwriteweb.com/enterprise/2010/10/google-apps-scripts.php

[57] S. King, "Google Restructures Its Principal AI Ethics Group," CyberEd.io. Accessed: Sep. 10, 2024. [Online]. Available: https://cybered.io/insights/google-restructures-its-principal-ai-ethics-group/

[58] P. Dave, "Google Splits Up a Key AI Ethics Watchdog," Wired, Jan. 31, 2024. Accessed: Sep. 10, 2024. [Online]. Available: https://www.wired.com/story/google-splits-up-responsible-innovation-ai-team/

[59] P. Dave and C. Haskins, "Google Lifts a Ban on Using Its AI for Weapons and Surveillance" Wired, Feb 4, 2025. [Online]. Available: https://www.wired.com/story/google-responsible-ai-principles

[60] J. S. Nelson, "The business ethics of Elon Musk, Tesla, Twitter and the tech industry," Harvard Law School. Accessed: Oct. 07, 2024. [Online]. Available: https://hls.harvard.edu/today/the-business-ethics-of-elon-musk-tesla-twitter-and-the-tech-industry/

[61] D. Lee, "The Moral Case for No Longer Engaging With Elon Musk's X," Bloomberg.com, Oct. 05, 2023. Accessed: Oct. 07, 2024. [Online]. Available: https://www.bloomberg.com/opinion/articles/2023-10-05/the-moral-case-for-no-longer-engaging-with-elon-musk-s-x

[62] "US Supreme Court won't hear Elon Musk dispute over SEC settlement," Reuters, Apr. 29, 2024. Accessed: Oct. 09, 2024. [Online]. Available: https://www.reuters.com/legal/us-supreme-court-wont-hear-elon-musk-dispute-over-sec-settlement-2024-04-29/

[63] S. Wong, F. Thorp V, R. Nobles, and L. Brown-Kaiser, "Elon Musk warns of 'civilizational risk' posed by AI in meeting with tech CEOs and senators," NBC News. Accessed: Oct. 07, 2024. [Online]. Available: https://www.nbcnews.com/politics/congress/big-tech-ceos-ai-meeting-senators-musk-zuckerberg-rcna104738

[64] xAI, "Acceptable Use Policy." Accessed: Oct. 07, 2024. [Online]. Available: https://x.ai/legal/enterprise/acceptable-use-policy

[65] B. Marr, "AI Gone Wild: How Grok-2 Is Pushing The Boundaries Of Ethics And Innovation," Forbes. Accessed: Oct. 04, 2024. [Online]. Available: https://www.forbes.com/sites/bernardmarr/2024/08/21/ai-gone-wild-how-grok-2-is-pushing-the-boundaries-of-ethics-and-innovation/

[66] K. Hammond, "Misinformation at Scale: Elon Musk's Grok and the Battle for Truth." Accessed: Oct. 07, 2024. [Online]. Available: https://casmi.northwestern.edu/news/articles/2024/misinformation-at-scale-elon-musks-grok-and-the-battle-for-truth

[67] B. Light and K. McGrath, "Ethics and social networking sites: a disclosive analysis of Facebook," Inf. Technol. People, vol. 23, no. 4, pp. 290–311, Jan. 2010, doi: 10.1108/09593841011087770.

[68] Meta, "Build Responsibly: The Meta Code of Conduct." Accessed: Oct. 22, 2024. [Online]. Available: https://scontent-bos5-1.xx.fbcdn.net/v/t39.8562-6/408715313_915324716774376_3554524540879262185_n.pdf?_nc_cat=111&ccb=1-7&_nc_sid=e280be&_nc_ohc=jdYaXyE833EQ7kNvgFQueMG&_nc_zt=14&_nc_ht=scontent-bos5-1.xx&_nc_gid=A4WUwuBipT1cITu1g4_yJyY&oh=00_AYCDLo3c8g1aplJvoraRtN57OE6Xq4TaNZ8go3FXCA6K_A&oe=671D8BB6

[69] B. Ortutay and H. Hadero, "Meta, TikTok and other social media CEOs testify in heated Senate hearing on child exploitation," AP News. Accessed: Oct. 22, 2024. [Online]. Available: https://apnews.com/article/meta-tiktok-snap-discord-zuckerberg-testify-senate-00754a6bea92aaad62585ed55f219932

[70] S. L. Harris, "FOCAL Programming Manual.pdf." Accessed: Oct. 22, 2024. [Online]. Available: http://www.bitsavers.org/www.computer.museum.uq.edu.au/pdf/DEC-08-AJAB-D%20PDP-8-I%20FOCAL%20Programming%20Manual.pdf

[71] D. Lauer, "Facebook's ethical failures are not accidental; they are part of the business model," AI Ethics, vol. 1, no. 4, pp. 395–403, Nov. 2021, doi: 10.1007/s43681-021-00068-x.

[72] L. H. Newman, "WhatsApp's New Privacy Policy Just Kicked In. Here's What You Need to Know," Wired, May 15, 2021. Accessed: Oct. 10, 2024. [Online]. Available: https://www.wired.com/story/whatsapp-privacy-policy-facebook-data-sharing/

[73] C. Hughes, "Anthropic Leads the Charge in AI Safety and Performance," Cloud Wars. Accessed: Oct. 04, 2024. [Online]. Available: https://cloudwars.com/ai/anthropic-leads-the-charge-in-ai-safety-and-performance/

[74] S. Moss, "Eleven OpenAI Employees Break Off to Establish Anthropic, Raise $124 Million | AI Business." Accessed: Oct. 04, 2024. [Online]. Available: https://aibusiness.com/verticals/eleven-openai-employees-break-off-to-establish-anthropic-raise-124m

[75] C. Duffy, "More OpenAI drama: Exec quits over concerns about focus on profit over safety | CNN Business," CNN. Accessed: Oct. 04, 2024. [Online]. Available: https://www.cnn.com/2024/05/17/tech/openai-exec-exits-safety-concerns/index.html

[76] C. Anil et al., "Many-shot Jailbreaking," Apr. 2024.

[77] "Building AI With a Conscience," The Atlantic. Accessed: Oct. 03, 2024. [Online]. Available: https://www.theatlantic.com/sponsored/google-2023/building-ai-with-a-conscience/3880/

[78] "Our Approach to User Safety | Anthropic Help Center." Accessed: Oct. 04, 2024. [Online]. Available: https://support.anthropic.com/en/articles/8106465-our-approach-to-user-safety

[79] "Claude's Constitution." Accessed: Oct. 04, 2024. [Online]. Available: https://www.anthropic.com/news/claudes-constitution

[80] R. Fabbro, "Google's multi-billion dollar relationship with Anthropic is under investigation," Quartz. Accessed: Oct. 04, 2024. [Online]. Available: https://qz.com/google-alphabet-anthropic-investment-investigation-1851608599

[81] G. Hammond, "AI start-up Anthropic accused of 'egregious' data scraping," Financial Times, Jul. 26, 2024. Accessed: Oct. 04, 2024. [Online]. Available: https://www.ft.com/content/07611b74-3d69-4579-9089-f2fc2af61baa

[82] S. Samuel, "It's practically impossible to run a big AI company ethically," Vox. Accessed: Oct. 04, 2024. [Online]. Available: https://www.vox.com/future-perfect/364384/its-practically-impossible-to-run-a-big-ai-company-ethically

[83] R. A. Spinello, "The case against Microsoft: An ethical perspective," Bus. Ethics Eur. Rev., vol. 12, no. 2, pp. 116–132, 2003, doi: 10.1111/1467-8608.00312.

[84] G. E. Dann and N. Haddow, "Just Doing Business or Doing Just Business: Google, Microsoft, Yahoo! and the Business of Censoring China's Internet," J. Bus. Ethics, vol. 79, no. 3, pp. 219–234, May 2008, doi: 10.1007/s10551-007-9373-9.

[85] Z. Schiffer and C. Newton, "Microsoft just laid off one of its responsible AI teams," Platformer. Accessed: Oct. 27, 2024. [Online]. Available: https://www.platformer.news/microsoft-just-laid-off-one-of-its/

[86] E. Ajao, "Microsoft whistleblower, OpenAI, the NYT, and ethical AI | TechTarget," Enterprise AI. Accessed: Oct. 27, 2024. [Online]. Available: https://www.techtarget.com/searchenterpriseai/news/366572699/Microsoft-whistleblower-OpenAI-the-NYT-and-ethical-AI

[87] S. Levy, "Trae Stephens Has Built AI Weapons and Worked for Donald Trump. As He Sees It, Jesus Would Approve | WIRED," Wired. Accessed: Oct. 12, 2024. [Online]. Available: https://www.wired.com/story/big-interview-trae-stephens-has-built-ai-weapons-and-worked-for-donald-trump-as-he-sees-it-jesus-would-approve/

[88] J. Ward and C. Sottile, "Inside Anduril, the startup that is building AI-powered military technology," NBC News. Accessed: Oct. 27, 2024. [Online]. Available: https://www.nbcnews.com/tech/security/inside-anduril-startup-building-ai-powered-military-technology-n1061771

[89] E. Spiers, "When Money Talks — And Says Horrible, Bigoted Things," There Is Only R. Accessed: Oct. 27, 2024. [Online]. Available: https://medium.com/there-is-only-r/when-money-talks-and-says-horrible-bigoted-things-e016b185592

[90] I. Fried, "AI firms treat any 'publicly available' data as fair game," Axios. Accessed: Oct. 01, 2024. [Online]. Available: https://www.axios.com/2024/04/05/open-ai-training-data-public-available-meaning

[91] S. Blake, "AI has taken over hundreds of jobs," Newsweek. Accessed: Oct. 01, 2024. [Online]. Available: https://www.newsweek.com/klarna-artificial-intelligence-tool-takes-700-jobs-1874002

[92] T. Germain, "AI took their jobs. Now they get paid to make it sound human." Accessed: Sep. 10, 2024. [Online]. Available: https://www.bbc.com/future/article/20240612-the-people-making-ai-sound-more-human

[93] B. Laker, "Adapting To AI: Interesting Insights From LinkedIn On The Job Market," Forbes. Accessed: Oct. 01, 2024. [Online]. Available: https://www.forbes.com/sites/benjaminlaker/2023/11/21/adapting-to-ai-interesting-insights-from-linkedin-on-the-job-market/

[94] J. Diamond, Collapse: How Societies Choose to Fail or Succeed: Revised Edition. Penguin, 2011.

[95] A. C. Kay et al., "Inequality, discrimination, and the power of the status quo: Direct evidence for a motivation to see the way things are as the way they should be," J. Pers. Soc. Psychol., vol. 97, no. 3, pp. 421–434, 2009, doi: 10.1037/a0015997.

[96] A. Potasznik, "Spokespeople or Gatekeepers? A Case Study of Teacher Engagement with the Data Privacy Policy Environment in U.S. Public Schools," Ph.D., University of Massachusetts Boston, United States -- Massachusetts, 2020. Accessed: Oct. 01, 2024. [Online]. Available: https://www.proquest.com/docview/2446721921/abstract?parentSessionId=dx1EQzjCFMNLeluLZ4frx74m89bP%2B2vsPOLPk1xDsTY%3D&sourcetype=Dissertations%20&%20Theses

[97] "US state-by-state AI legislation snapshot," BCLP - Bryan Cave Leighton Paisner - US state-by-state AI legislation snapshot. Accessed: Oct. 01, 2024. [Online]. Available: https://www.bclplaw.com/en-US/events-insights-news/us-state-by-state-artificial-intelligence-legislation-snapshot.html

[98] "Blueprint for an AI Bill of Rights | OSTP," The White House. Accessed: Oct. 01, 2024. [Online]. Available: https://www.whitehouse.gov/ostp/ai-bill-of-rights/

[99] S. Tannenbaum, "Black-owned businesses saw a surge of support after George Floyd's murder. Since then? Not so much. - The Boston Globe," BostonGlobe.com. Accessed: Oct. 05, 2024. [Online]. Available: https://www.bostonglobe.com/2024/10/05/business/black-owned-business-boston-george-floyd/

[100] R. Ravaglia, "Google Gemini Provides Generative AI Tutoring With Moral Exhortation," Forbes. Accessed: Oct. 04, 2024. [Online]. Available:https://www.forbes.com/sites/rayravaglia/2024/03/03/google-gemini-provides-generative-ai-tutoring-with-moral-exhortation/

[101] M. Isaac, "Meta Permits Its A.I. Models to Be Used for U.S. Military Purposes," *The New York Times*, Nov. 05, 2024. Accessed: Nov. 05, 2024. [Online]. Available: https://www.nytimes.com/2024/11/04/technology/meta-ai-military.html

[102] J. Pomfret and J. Pang, "Exclusive: Chinese Researchers develop AI model for military use on back of Meta's Llama," Reuters, Nov. 01, 2024. [Online]. Available: https://www.reuters.com/technology/artificial-intelligence/chinese-researchers-develop-ai-model-military-use-back-metas-llama-2024-11-01/

[103] L. Ropek, "OpenAI Continues Its Mission of 'Ethical' AI by Partnering With a Killer Robot Company," Gizmodo. Accessed: Apr. 06, 2025. [Online]. Available: https://gizmodo.com/openai-continues-its-mission-of-ethical-ai-by-partnering-with-a-killer-robot-company-2000534515