

**The
Intercept_**

THE INVENTION OF “ETHICAL AI”

How Big Tech Manipulates Academia to Avoid Regulation

Illustration: Yoshi Sodeoka for The Intercept



Rodrigo Ochigame

December 20 2019, 7:19 p.m.

The irony of the ethical scandal enveloping Joichi Ito, the former director of the MIT Media Lab, is that he used to lead academic initiatives on ethics. After the revelation of his financial ties to Jeffrey Epstein, the financier charged with sex trafficking underage girls as young as 14, Ito resigned from multiple roles at MIT, a visiting professorship at Harvard Law School, and the boards of the John D. and Catherine T. MacArthur Foundation, the John S. and James L. Knight Foundation, and the New York Times Company.

Many spectators are puzzled by Ito’s influential role as an ethicist of artificial intelligence. Indeed, his initiatives were crucial in establishing the discourse of “ethical AI” that is now ubiquitous in academia

and in the mainstream press. In 2016, then-President Barack Obama described him as an “expert” on AI and ethics. Since 2017, Ito financed many projects through the \$27 million [Ethics and Governance of AI Fund](#), an initiative anchored by the MIT Media Lab and the Berkman Klein Center for Internet and Society at Harvard University. What was all the talk of “ethics” really about?

For 14 months, I worked as a graduate student researcher in Ito’s group on AI ethics at the Media Lab. I stopped on August 15, immediately after Ito [published](#) his initial “apology” regarding his ties to Epstein, in which he acknowledged accepting money from the financier both for the Media Lab and for Ito’s outside venture funds. Ito did not disclose that Epstein had, at the time this money changed hands, already pleaded guilty to a child prostitution charge in Florida, or that Ito took numerous steps to hide Epstein’s name from official records, as The New Yorker [later revealed](#).

Inspired by whistleblower Signe Swenson and others who have spoken out, I have decided to report what I came to learn regarding Ito’s role in shaping the field of AI ethics, since this is a matter of public concern. The emergence of this field is a recent phenomenon, as past AI researchers had been largely uninterested in the study of ethics. A former Media Lab colleague recalls that Marvin Minsky, the deceased AI pioneer at MIT, used to say that “an ethicist is someone who has a problem with whatever you have in your mind.” (In recently unsealed court filings, victim Virginia Roberts

The discourse of “ethical AI” was aligned strategically with a Silicon Valley effort seeking to avoid legally enforceable restrictions of controversial technologies.

Giuffre testified that Epstein directed her to have sex with Minsky.) Why, then, did AI researchers suddenly start talking about ethics?

At the Media Lab, I learned that the discourse of “ethical AI,” championed substantially by Ito, was aligned strategically with a Silicon Valley effort seeking to avoid legally enforceable restrictions of controversial technologies. A key group behind this effort, with the lab as a member, made policy recommendations in California that contradicted the conclusions of research I conducted with several lab colleagues, research that led us to oppose the use of computer algorithms in deciding whether to jail people pending trial. Ito himself would eventually complain, in private meetings with financial and tech executives, that the group’s recommendations amounted to “whitewashing” a thorny ethical issue. “They water down stuff we try to say to prevent the use of algorithms that don’t seem to work well” in detention decisions, he confided to one billionaire.

I also watched MIT help the U.S. military brush aside the moral complexities of drone warfare, hosting a superficial talk on AI and ethics by Henry Kissinger, the former secretary of state and notorious war criminal, and giving input on the U.S. Department of Defense’s “AI Ethics Principles” for warfare, which embraced “permissibly biased” algorithms and which avoided using the word “fairness” because the Pentagon believes “that fights should not be fair.”

Ito did not respond to requests for comment.



Joichi Ito, then-director of MIT Media Lab, speaks during a press conference in Tokyo on July 8, 2016. Photo: Akio Kon/Bloomberg/Getty Images

MIT lent credibility to the idea that big tech could police its own use of artificial intelligence at a time when the industry faced increasing criticism and calls for legal regulation. Just in 2018, there were several controversies: Facebook's breach of private data on more than 50 million users to a political marketing firm hired by Donald Trump's presidential campaign, revealed in March 2018; Google's contract with the Pentagon for computer vision software to be used in combat zones, revealed that same month; Amazon's sale of facial recognition technology to police departments, revealed in May; Microsoft's contract with the U.S. Immigration and Customs Enforcement revealed in June; and IBM's secret collaboration with the New York Police Department for facial recognition and racial classification in video surveillance footage, revealed in September. Under the slogan #TechWontBuildIt, thousands of workers at these firms have organized protests and circulated petitions against such contracts. From #NoTechForICE to #Data4BlackLives, several grassroots campaigns have demanded legal

restrictions of some uses of computational technologies (e.g., forbidding the use of facial recognition by police).

Meanwhile, corporations have tried to shift the discussion to focus on voluntary “ethical principles,” “responsible practices,” and technical adjustments or “safeguards” framed in terms of “bias” and “fairness” (e.g., requiring or encouraging police to adopt “unbiased” or “fair” facial recognition). In January 2018, Microsoft published its “ethical principles” for AI, starting with “fairness.” In May, Facebook announced its “commitment to the ethical development and deployment of AI” and a tool to “search for bias” called “Fairness Flow.” In June, Google published its “responsible practices” for AI research and development. In September, IBM announced a tool called “AI Fairness 360,” designed to “check for unwanted bias in datasets and machine learning models.” In January 2019, Facebook granted \$7.5 million for the creation of an AI ethics center in Munich, Germany. In March, Amazon co-sponsored a \$20 million program on “fairness in AI” with the U.S. National Science Foundation. In April, Google canceled its AI ethics council after [backlash](#) over the selection of Kay Coles James, the vocally anti-trans president of the right-wing Heritage Foundation. These corporate initiatives frequently cited academic research that Ito had supported, at least partially, through the MIT-Harvard fund.

To characterize the corporate agenda, it is helpful to distinguish between three kinds of regulatory possibilities for a given technology: (1) no legal regulation at all, leaving “ethical principles” and “responsible practices” as merely voluntary; (2) moderate legal regulation encouraging or requiring technical adjustments that do not conflict significantly with profits; or (3) restrictive legal regulation curbing or banning deployment of the technology. Unsurprisingly, the tech industry tends to support the first two and oppose the last. The corporate-sponsored discourse of “ethical AI” enables precisely this position. Consider the case of facial recognition. This year, the municipal legislatures of San Francisco, Oakland, and Berkeley – all in Califor-

nia – plus Somerville, Massachusetts, have passed strict bans on facial recognition technology. Meanwhile, Microsoft has lobbied in favor of less restrictive legislation, requiring technical adjustments such as tests for “bias,” most notably in Washington state. Some big firms may even prefer this kind of mild legal regulation over a complete lack thereof, since larger firms can more easily invest in specialized teams to develop systems that comply with regulatory requirements.

Thus, Silicon Valley’s vigorous promotion of “ethical AI” has constituted a strategic lobbying effort, one that has enrolled academia to legitimize itself. Ito played a key role in this corporate-academic fraternizing, meeting regularly with tech executives. The MIT-Harvard fund’s initial director was the former “global public policy lead” for AI at Google. Through the fund, Ito and his associates sponsored many projects, including the creation of a prominent conference on “Fairness, Accountability, and Transparency” in computer science; other sponsors of the conference included Google, Facebook, and Microsoft.

Join Our Newsletter

Original reporting. Fearless journalism. Delivered to you.

I'm in →

Although the Silicon Valley lobbying effort has consolidated academic interest in “ethical AI” and “fair algorithms” since 2016, a handful of papers on these topics had appeared in earlier years, even if framed differently. For example, Microsoft computer scientists published the [paper](#) that arguably inaugurated the field of “algorithmic fairness” in 2012. In 2016, the paper’s lead author, Cynthia Dwork, became a professor of computer science at Harvard, with simultaneous positions at its law school and at Microsoft. When I took her Harvard course on the mathematical foundations of cryptography and statistics in 2017, I interviewed her and asked how she became interested in researching algorithmic definitions of fairness. In her account, she had long been

personally concerned with the issue of discriminatory advertising, but Microsoft managers encouraged her to pursue this line of work because the firm was developing a new system of online advertising, and it would be economically advantageous to provide a service “free of regulatory problems.” (To be fair, I believe that Dwork’s personal intentions were honest despite the corporate capture of her ideas. Microsoft declined to comment for this article.)

After the initial steps by MIT and Harvard, many other universities and new institutes received money from the tech industry to work on AI ethics. Most such organizations are also headed by current or former executives of tech firms. For example, the Data & Society Research Institute is directed by a Microsoft researcher and initially funded by a Microsoft grant; New York University’s AI Now Institute was co-founded by another Microsoft researcher and partially funded by Microsoft, Google, and DeepMind; the Stanford Institute for Human-Centered AI is co-directed by a former vice president of Google; University of California, Berkeley’s Division of Data Sciences is headed by a Microsoft veteran; and the MIT Schwarzman College of Computing is headed by a board member of Amazon. During my time at the Media Lab, Ito maintained frequent contact with the executives and planners of all these organizations.

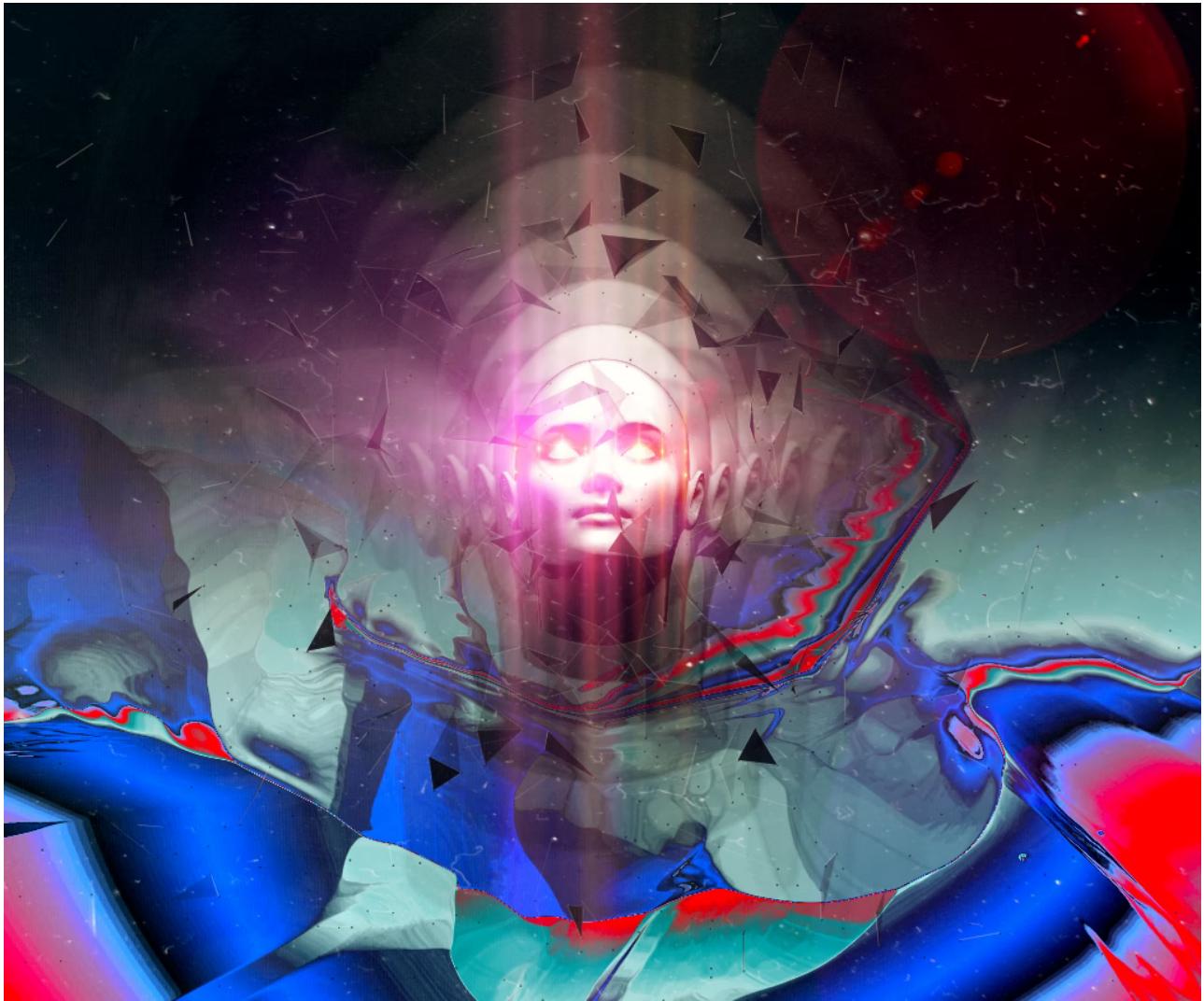


Illustration: Yoshi Sodeoka for The Intercept

Big tech money and direction proved incompatible with an honest exploration of ethics, at least judging from my experience with the “Partnership on AI to Benefit People and Society,” a group founded by Microsoft, Google/DeepMind, Facebook, IBM, and Amazon in 2016. PAI, of which the Media Lab is a member, defines itself as a “multistakeholder body” and claims it is “not a lobbying organization.” In an April 2018 hearing at the U.S. House Committee on Oversight and Government Reform, the Partnership’s executive director claimed that the organization is merely “*a resource to policymakers* – for instance, in conducting research that informs AI best practices and exploring the societal consequences of certain AI systems, as well as policies around the development and use of AI systems.”

But even if the Partnership's activities may not meet the legal threshold requiring registration as lobbyists – for example, by seeking to directly affect the votes of individual elected officials – the partnership has certainly sought to influence legislation. For example, in November 2018, the Partnership staff asked academic members to contribute to a collective statement to the Judicial Council of California regarding a Senate bill on penal reform (S.B. 10). The bill, in the course of eliminating cash bail, expanded the use of algorithmic risk assessment in pretrial decision making, and required the Judicial Council to “address the identification and mitigation of any implicit bias in assessment instruments.” The Partnership staff wrote, “we believe there is room to impact this legislation (and CJS [criminal justice system] applications more broadly).”

In December 2018, three Media Lab colleagues and I raised serious objections to the Partnership's efforts to influence legislation. We observed that the Partnership's policy recommendations aligned consistently with the corporate agenda. In the penal case, our research led us to strongly oppose the adoption of risk assessment tools, and to reject the proposed technical adjustments that would supposedly render them “unbiased” or “fair.” But the Partnership's draft statement seemed, as a colleague put it in an internal email to Ito and others, to “validate the use of RA [risk assessment] by emphasizing the issue as a technical one that can therefore be solved with better data sets, etc.” A second colleague agreed that the “PAI statement is weak and risks doing exactly what we've been warning against re: the risk of legitimization via these industry led regulatory efforts.” A third colleague wrote, “So far as the criminal justice work is concerned, what PAI is doing in this realm is quite alarming and also in my opinion seriously misguided. I agree with Rodrigo that PAI's association with ACLU, MIT and other academic / non-profit institutions practically ends up serving a legitimating function. Neither ACLU nor MIT nor any non-profit has any power in PAI.”

Worse, there seemed to be a mismatch between the Partnership's recommendations and the efforts of a grassroots coalition of organizations fighting jail expansion, including the movement Black Lives Matter, the prison abolitionist group Critical Resistance (where I have volunteered), and the undocumented and queer/trans youth-led Immigrant Youth Coalition. The grassroots coalition argued, "The notion that any risk assessment instrument can account for bias ignores the racial disparities in current and past policing practices." There are abundant theoretical and empirical reasons to support this claim, since risk assessments are typically based on data of arrests, convictions, or incarcerations, all of which are poor proxies for individual behaviors or predispositions. The coalition continued, "Ultimately, risk-assessment tools create a feedback-loop of racial profiling, pre-trial detention and conviction. A person's freedom should not be reduced to an algorithm." By contrast, the Partnership's statement focused on "minimum requirements for responsible deployment," spanning such topics as "validity and data sampling bias, bias in statistical predictions; choice of the appropriate targets for prediction; human-computer interaction questions; user training; policy and governance; transparency and review; reproducibility, process, and recordkeeping; and post-deployment evaluation."

To be sure, the Partnership staff did respond to criticism of the draft by noting in the final version of the statement that "within PAI's membership and the wider AI community, many experts further suggest that individuals can never justly be detained on the basis of their risk assessment score alone, without an individualized hearing." This meek concession – admitting that it might not be time to start imprisoning people based strictly on software, without input from a judge or any other "individualized" judicial process – was easier to make because none of the major firms in the Partnership sell risk assessment tools for pretrial decision-making; not only is the technology too controversial but also the market is too small. (Facial recognition technology, on the other hand, has a much larger market in which Microsoft, Google, Facebook, IBM, and Amazon all operate.)

In December 2018, my colleagues and I urged Ito to quit the Partnership. I argued, “If academic and nonprofit organizations want to make a difference, the only viable strategy is to quit PAI, make a public statement, and form a counter alliance.” Then a colleague proposed, “there are many other organizations which are doing much more substantial and transformative work in this area of predictive analytics in criminal justice – what would it look like to take the money we currently allocate in supporting PAI in order to support their work?” We believed Ito had enough autonomy to do so because the MIT-Harvard fund was supported largely by the Knight Foundation, even though most of the money came from tech investors Pierre Omidyar, founder of eBay, via the Omidyar Network, and Reid Hoffman, co-founder of LinkedIn and Microsoft board member. I wrote, “If tens of millions of dollars from nonprofit foundations and individual donors are not enough to allow us to take a bold position and join the right side, I don’t know what would be.” (Omidyar funds The Intercept.)

It is strange that Ito, with no formal training, became positioned as an “expert” on AI ethics, a field that barely existed before 2017.

Ito did acknowledge the problem. He had just received a message from David M. Siegel, co-chair of the hedge fund Two Sigma and member of the MIT Corporation. Siegel proposed a self-regulatory structure for “search and social media” firms in Silicon Valley, modeled after the Financial Industry Reg-

ulatory Authority, or FINRA, a private corporation that serves as a self-regulatory organization for securities firms on Wall Street. Ito responded to Siegel’s proposal, “I don’t feel civil society is well represented in the industry groups. We’ve been participating in Partnership in AI and they water down stuff we try to say to prevent the use

of algorithms that don't seem to work well like risk scores for pre-trial bail. I think that with personal data and social media, I have concerns with self-regulation. For example, a full blown genocide [of the Rohingya, a mostly Muslim minority group in Myanmar] happened using What's App and Facebook knew it was happening." (Facebook has **admitted** that its platform was used to incite violence in Myanmar; news reports have documented how content on the Facebook platform **facilitated** a genocide in the country despite **repeated warnings** to Facebook executives from human rights activists and researchers. Facebook texting service WhatsApp made it harder for its users to forward messages after WhatsApp was **reportedly** used to spread misinformation during elections in India.)

But the corporate-academic alliances were too robust and convenient. The Media Lab remained in the Partnership, and Ito continued to fraternize with Silicon Valley and Wall Street executives and investors. Ito described Siegel, a billionaire, as a "potential funder." With such people, I saw Ito routinely express moral concerns about their businesses – but in a friendly manner, as he was simultaneously asking them for money, whether for MIT or his own venture capital funds. For corporate-academic "ethicists," amicable criticism can serve as leverage for entering into business relationships. Siegel replied to Ito, "I would be pleased to speak more on this topic with you. Finra is not an industry group. It's just paid for by industry. I will explain more when we meet. I agree with your concerns."

In private meetings, Ito and tech executives discussed the corporate lobby quite frankly. In January, my colleagues and I joined a meeting with Mustafa Suleyman, founding co-chair of the Partnership and co-founder of DeepMind, an AI startup acquired by Google for about \$500 million in 2014. In the meeting, Ito and Suleyman discussed how the promotion of "AI ethics" had become a "whitewashing" effort, although they claimed their initial intentions had been nobler. In a message to plan the meeting, Ito wrote to my colleagues and me, "I do know, however, from speaking to Mustafa when he was setting up

PAI that he was meaning for the group to be much more substantive and not just ‘white washing.’ I think it’s just taking the trajectory that these things take.” (Suleyman did not respond to requests for comment.)

Regardless of individual actors’ intentions, the corporate lobby’s effort to shape academic research was extremely successful. There is now an enormous amount of work under the rubric of “AI ethics.” To be fair, some of the research is useful and nuanced, especially in the humanities and social sciences. But the majority of well-funded work on “ethical AI” is aligned with the tech lobby’s agenda: to voluntarily or moderately adjust, rather than legally restrict, the deployment of controversial technologies. How did five corporations, using only a small fraction of their budgets, manage to influence and frame so much academic activity, in so many disciplines, so quickly? It is strange that Ito, with no formal training, became positioned as an “expert” on AI ethics, a field that barely existed before 2017. But it is even stranger that two years later, respected scholars in established disciplines have to demonstrate their relevance to a field conjured by a corporate lobby.



Former Google CEO Eric Schmidt, now chair of the Department of Defense's Defense Innovation Board, takes his seat for the House Armed Services Committee hearing on "Promoting DOD's Culture of Innovation" on April 17, 2018. Photo: Bill Clark/CQ Roll Call/Getty Images

The field has also become relevant to the U.S. military, not only in official responses to moral concerns about technologies of targeted killing but also in disputes among Silicon Valley firms over lucrative military contracts. On November 1, the Department of Defense's innovation board published its recommendations for "AI Ethics Principles." The board is chaired by Eric Schmidt, who was the executive chair of Alphabet, Google's parent company, when Obama's defense secretary Ashton B. Carter established the board and appointed him in 2016. According to ProPublica, "Schmidt's influence, already strong under Carter, only grew when [James] Mattis arrived as [Trump's] defense secretary." The board includes multiple executives from Google, Microsoft, and Facebook, raising controversies regarding conflicts of interest. A Pentagon employee responsible for policing conflicts of interest was removed from the innovation board after she challenged "the Pentagon's cozy relationship not only with [Amazon CEO Jeff] Bezos, but with Google's Eric Schmidt." This relationship is potentially lucrative for big tech firms: The AI ethics recommendations

appeared less than a week after the Pentagon awarded a \$10 billion cloud-computing contract to Microsoft, which is being legally challenged by Amazon.

The recommendations seek to compel the Pentagon to increase military investments in AI and to adopt “ethical AI” systems such as those developed and sold by Silicon Valley firms. The innovation board calls the Pentagon a “deeply ethical organization” and offers to extend its “existing ethics framework” to AI. To this end, the board cites the AI ethics research groups at Google, Microsoft, and IBM, as well as academics sponsored by the MIT-Harvard fund. However, there are caveats. For example, the board notes that although “the term ‘fairness’ is often cited in the AI community,” the recommendations avoid this term because of “the DoD mantra that fights should not be fair, as DoD aims to create the conditions to maintain an unfair advantage over any potential adversaries.” Thus, “some applications will be permissibly and justifiably biased,” specifically “to target certain adversarial combatants more successfully.” The Pentagon’s conception of AI ethics forecloses many important possibilities for moral deliberation, such as the prohibition of drones for targeted killing.

The corporate, academic, and military proponents of “ethical AI” have collaborated closely for mutual benefit. For example, Ito told me that he informally advised Schmidt on which academic AI ethicists

The majority of well-funded work on “ethical AI” is aligned with the tech lobby’s agenda: to voluntarily or moderately adjust, rather than legally restrict, the deployment of controversial technologies.

Schmidt's private foundation should fund. Once, Ito even asked me for second-order advice on whether Schmidt should fund a certain professor who, like Ito, later served as an "expert consultant" to the Pentagon's innovation board. In February, Ito joined Carter at a panel titled "Computing for the People: Ethics and AI," which also included current and former executives of Microsoft and Google. The panel was part of the inaugural celebration of MIT's \$1 billion college dedicated to AI. Other speakers at the celebration included Schmidt on "Computing for the Marketplace," Siegel on "How I Learned to Stop Worrying and Love Algorithms," and Henry Kissinger on "How the Enlightenment Ends." As Kissinger declared the possibility of "a world relying on machines powered by data and algorithms and ungoverned by ethical or philosophical norms," a [protest outside the MIT auditorium](#) called attention to Kissinger's war crimes in Vietnam, Cambodia, and Laos, as well as his support of war crimes elsewhere. In the age of automated targeting, what atrocities will the U.S. military justify as governed by "ethical" norms or as executed by machines beyond the scope of human agency and culpability?

No defensible claim to "ethics" can sidestep the urgency of legally enforceable restrictions to the deployment of technologies of mass surveillance and systemic violence. Until such restrictions exist, moral and political deliberation about computing will remain subsidiary to the profit-making imperative expressed by the Media Lab's motto, "Deploy or Die." While some deploy, even if ostensibly "ethically," others die.

WAIT! BEFORE YOU GO on about your day, ask yourself: How likely is it that the story you just read would have been produced by a different news outlet if The Intercept hadn't done it?

Consider what the world of media would look like without The Intercept. Who would hold party elites accountable to the values they proclaim to have? How many covert wars, miscarriages of justice, and dystopian technologies would remain hidden if our reporters weren't on the beat?

The kind of reporting we do is essential to democracy, but it is not easy, cheap, or profitable. The Intercept is an independent nonprofit news outlet. We don't have ads, so we depend on our members – 35,000 and counting – to help us hold the powerful to account. Joining is simple and doesn't need to cost a lot: You can become a sustaining member for as little as \$3 or \$5 a month. That's all it takes to support the journalism you rely on.

We're independent of corporate interests. Will you help us?

\$15

\$25

\$50

\$100

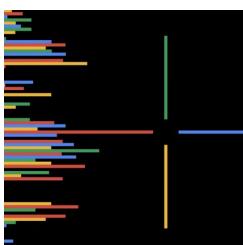
ONE-TIME

MONTHLY

Become a Member



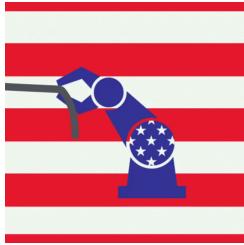
RELATED



Google Continues Investments in Military and Police AI Technology Through Venture Capital Arm



Facebook Uses Artificial Intelligence to Predict Your Future Actions for Advertisers, Says Confidential Document



Why an “AI Race” Between the U.S. and China Is a Terrible, Terrible Idea



Amazon Promises “Unwavering” Commitment to Police, Military Clients Using AI Technology

LATEST STORIES



Congressional Democrats Join Republicans to Undermine Biden Administration’s Surprise Medical Billing Rule

Austin Ahlman – Jan. 17

Worried an aggressive new rule could cut into providers’ earnings, key members of Congress including Massachusetts Representative Richard Neal are helping private industry weaken the administration’s position in federal court.



Voices

2021’s Inflation Proves How Incredibly Valuable Social Security



Is for Americans

[Jon Schwarz](#) – Jan. 16

Social Security is inflation-adjusted, so benefits are going up 5.9 percent to keep the purchasing power of retirees the same.



Making a Killing UAE Adviser Illegally Funneled Foreign Cash Into Hillary Clinton's 2016 Campaign

[Matthew Cole](#) – Jan. 16

George Nader also cultivated key Trump advisers on behalf of his Gulf clients, prosecutors say.



JOIN THE CONVERSATION

Fetching more

