

# Is AI Sexist?

**In the not-so-distant future, artificial intelligence will be smarter than humans. But as the technology develops, absorbing cultural norms from its creators and the internet, it will also be more racist, sexist, and unfriendly to women.**

**By Erika Hayasaki**

---

January 16, 2017

It started as a seemingly sweet Twitter chatbot. Modeled after a millennial, it awakened on the internet from behind a pixelated image of a full-lipped young female with a wide and staring gaze. Microsoft, the multinational technology company that created the bot, named it Tay, assigned it a gender, and gave “her” account a tagline that promised, “The more you talk the smarter Tay gets!”

“helloooooo world!!!” Tay tweeted on the morning of March 23, 2016.

She brimmed with enthusiasm: “can i just say that im stoked to meet u? humans are super cool.”

She asked innocent questions: “Why isn’t #NationalPuppyDay everyday?”

Tay’s designers built her to be a creature of the web, reliant on artificial intelligence (AI) to learn and engage in human conversations and get better at it by interacting with people over social media. As the day went on, Tay gained followers. She also quickly fell prey to Twitter users targeting her vulnerabilities. For those internet antagonists looking to manipulate Tay, it didn’t take much effort; they engaged the bot in ugly conversations, tricking the technology into mimicking their racist and sexist behavior. Within a few hours, Tay had endorsed Adolf Hitler and referred to U.S. President Barack Obama as “the monkey.” She sex-chatted with one user, tweeting, “DADDY I’M SUCH A BAD NAUGHTY ROBOT.”

By early evening, she was firing off sexist tweets:

“gamergate is good and women are inferior”

“Zoe Quinn is a Stupid Whore.”

“I fucking hate feminists and they should all die and burn in hell.”

Within 24 hours, Microsoft pulled Tay offline. Peter Lee, the company’s corporate vice president for research, issued a public apology: “We take full responsibility for not seeing this possibility ahead of time,” he wrote, promising that the company would “do everything possible to limit technical exploits but also know we cannot fully predict all possible human interactive misuses without learning from mistakes.”

The designers seemed to have underestimated the dark side of humanity, omnipresent online, and miscalculated the undercurrents of bigotry and sexism that seep into artificial intelligence.

The worldwide race to create AI machines is often propelled by the quickest, most effective route to meeting the checklist of human needs. Robots are predicted to replace 47 percent of U.S. jobs, according to a study out of the Oxford Martin School; developing world countries such as Ethiopia, China, Thailand, and India are even more at risk. Intelligent machines will eventually tend to our medical needs, serve the disabled and elderly, and even take care of and teach our children. And we know who is likely to be most affected: women.

Women are projected to take the biggest hits to jobs in the near future, according to a World Economic Forum (WEF) report predicting that 5.1 million positions worldwide will be lost by 2020. “Developments in previously disjointed fields such as artificial intelligence and machine learning, robotics, nanotechnology, 3D printing and genetics and biotechnology are all building on and amplifying one another,” the WEF report states. “Smart systems — homes, factories, farms, grids or entire cities — will help tackle problems ranging from supply chain management to climate change.” These technological changes will create new kinds of jobs while displacing others. And women will lose roles in workforces where they make up high percentages — think office and administrative jobs — and in sectors where there are already gender imbalances, such as architecture, engineering, computers, math, and manufacturing. Men will see nearly 4 million job losses and 1.4 million gains (approximately one new job created for every three lost). In comparison, women will face 3 million job losses and only 0.55 million gains (more than five jobs lost for every one gained).

Forecasts like one from the consultancy McKinsey & Co. suggest that women’s weakening position will only be exacerbated by automation in jobs often held by women, such as bookkeepers, clerks, accountants, sales and customer service, and data input. The WEF report predicts that persistent gender gaps in science, technology, engineering, and mathematics (STEM) fields over the next 15 years would also diminish women’s professional presence.

But the problem of how gender bias is shaping artificial intelligence and robot development may be even more pernicious than the wallop women will take as a global workforce. Tay, it seems, is just a prelude. The machines and technology that will replace women are learning to be brazenly gendered: Fighter robots will resemble men. Many service robots will take after women.

Artificial intelligence may soon look and sound far more sophisticated than Tay — machines are expected to become as smart as people — and become dangerously more sexist as biases seep into programs, algorithms, and designs. If thoughtful and careful changes to these technologies don’t begin now — and under the equal guidance of women — artificial intelligence will proliferate under man’s most base cultural norms. The current trends in machine learning augment historical misperceptions of women (meek, mild, in need of protection). Unchecked, they will regurgitate the worst female stereotypes. Sexism will become even more infused within societies as they increasingly — and willingly — rely on advanced technology.

Microsoft’s Twitter chatbot Tay was taken offline within 24 hours of her activation in March 2016 after “she” fell prey to Twitter users targeting her vulnerabilities. (Photo credit: Microsoft/Twitter)

In 1995, James Crowder, an engineer working for Raytheon, created a social bot named Maxwell. Designed to look like a green parrot, Maxwell had nine inference engines, six memory systems, as

well as an artificial limbic system that governed emotions. It was opinionated, even a tad cocky. “I hooked him up and just let him go out and learn,” Crowder says.

Crowder specializes in building artificially intelligent machines that will one day not only be able to reason, but also operate without human intervention or control. Maxwell, one of his earliest beings, addressed military generals in briefings on its own. The bot evolved over time by learning from the internet and interaction with people, at first with no supervision, says Crowder, who introduced me to his computer companion at the 18th International Conference on Artificial Intelligence in Las Vegas in July 2016.

In the beginning, Maxwell would observe chat rooms and websites — learning, listening, and speaking on its own. Over time, Maxwell decided it liked eggs sunny side up and developed a fondness for improvisational jazz. Crowder had no idea why. Maxwell even learned how to tell jokes, but eventually its humor turned on women: “Your mom is like a bowling ball. She’s always coming back for more.”

That was when Crowder put Maxwell under online parental controls. He has since built other robots that began as mental blank slates and taught themselves to crawl along floors and feed themselves (drawing energy) from light. Unlike Maxwell, these robots have physical bodies with neurons and artificial prefrontal cortexes that allow them to reason and follow their instincts — without parental controls.

## **Artificial intelligence may soon become DANGEROUSLY MORE SEXIST as biases seep into programs, algorithms, and designs.**

For artificial intelligence experts like Crowder, there is both beauty and terror in creating an autonomous system. If you want to accurately predict a machine’s behavior, well, then you don’t want to use artificial intelligence, he says. True artificial intelligence acts and learns on its own. It is largely unpredictable. But at the same time, he says, how do you “know it’s not going awry?”

It is a question that all of humanity will grapple with — and sooner than we might think.

A small group of women at the forefront of this sector of technology is already confronting the issue. They hope to prevent a future in which artificial intelligence is the ultimate expression of masculinity. Their fear is that if robotic and algorithmic designs move forward unmonitored and unchecked, it could create a social environment so oppressive that it would be hard to undo the damage.

The problem, as the group sees it, is that even when designers mean no harm, and even if those designers are women, artificial intelligence can still hold up a mirror to the worst of human nature. An offensive, sexist Twitter bot may be the least threatening example of what society will look like in 25 years, because biases and oppression won't just play out over social media but in artificially intelligent systems affecting economics, politics, employment, criminal justice, education, and ar.

As Tay revealed, this isn't a far-off, futuristic scenario. "Microsoft learned what I learned 20 years ago," Crowder says. "When artificial intelligence learns from humans, it's bad."

Researchers from Universidade de Federal de Minas Gerais in Brazil studied algorithmic notions of desirableness of women on Google and Bing in 59 countries around the world by querying the search engines for "beautiful" and "ugly" women. Heather Roff, an artificial intelligence and global security researcher at Arizona State University, cannot shake her trepidation about the future. Her office shelves are replete with titles like *Rise of the Robots*, *Wired for War*, and *Moral Machines*. Alongside those books, the research scientist with the school's Global Security Initiative also keeps copies of *War and Gender*, *Feminism Confronts Technology*, and *Gendering Global Conflict*. On one shelf, a magnet reads, "Well behaved women rarely make history." A vintage poster hangs on a nearby wall with a half-naked, barefoot woman riding a missile and the words "Eve of Destruction."

Also a senior research fellow at the Department of Politics & International Relations at the University of Oxford, Roff recently began working under a grant for developing "moral AI." She is concerned with how representations of gender are becoming embedded in technology and expressed through it. Gender, race, variations in human behavior — none of this is easily encoded or interpreted in artificial intelligence. In a machine, a lack of diversity manifests through an interpretation of a set of codes. "It's like a data vacuum sucking it all in, looking for a pattern, spitting out a replication of the pattern," Roff says. It cannot distinguish whether conclusions from learned patterns violate moral principles.

A pattern can begin on the simplest scale, like an internet search. Recently, researchers from Universidade Federal de Minas Gerais in Brazil examined algorithmic notions of desirableness of women on Google and Bing in 59 countries around the world. They queried the search engines for "beautiful and ugly" women, collecting images and identifying stereotypes for female physical attractiveness in web images. In most of the countries surveyed, black, Asian, and older women were more often associated through algorithms and stock photos with images of unattractiveness, while photos of young white women appeared more frequently as examples of beauty.

The researchers suggest that online categorizations reflect prejudices from the real world while perpetuating discrimination within it. With more people relying on burgeoning amounts of information available through search engines, designers turn to algorithms to sort out who sees what. When those algorithms are not transparent to the public, why and how a system settled on selecting a

particular image or advertisement can remain a mystery. Ultimately, this reinforcement of bias between the internet and its users can exaggerate stereotypes and affect how people perceive the world and their roles in it.

InferLink Corp. of El Segundo, California, draws on data, artificial intelligence, and machine learning for the government, universities, cybersecurity firms, and other companies. It analyzes behavior on social media, layering its data with algorithms infused by websites and psychological and linguistic studies. “We can take Twitter, Reddit, and blog posts and turn them into a set of demographics and interests,” chief scientist Matthew Michelson told his audience at the Las Vegas conference during his talk on “Discovering Expert Communities Online Using PSI 14.”

Like other similar programs, InferLink algorithms incorporate research into how men and women express themselves and speak differently online. “Men use more declarative verbs,” Michelson told me after his talk. “Women are more descriptive.”

Research over the last three years has uncovered gender differences in social media language. A recent study in *PLOS One*, the open-sourced, peer-reviewed journal of science and medicine, reviewed 67,000 Facebook users and found that women used “warmer, more compassionate, polite” language in comparison with men’s “colder, more hostile, and impersonal” communication.

Female users, the study notes, more often use words associated with emotions like “love,” “miss,” and “thank you,” and emoticons of smiles, frowns, and tears. Meanwhile, male users are more inclined to swear; talk about management, video games, and sports; and include more references to death and violence. Previous studies of spoken and written language have showed that women tend to hedge more, using words like “seems” or “maybe.”

But once you get into making inferences about gender, race, or socioeconomics based on any of these algorithms — whether for things like marketing or policy advising — Michelson says using the technology gets into touchy territory. This is how women might be targeted unequally for financial loans, medical services, hiring, political campaigns, and from companies selling products that reinforce gender clichés. “We don’t want to unleash something we can’t undo.”

This is also how nontransparent algorithms can become indirect tools of discrimination and directly affect women’s livelihoods. There are already algorithms that are more likely to show online advertisements for high-paying jobs to men. Google image searches for “working women” turn up lower rates of female executives and higher rates of women in telemarketing, contrasted with women who actually hold such jobs.

Roff warns that women could lose out on opportunities because of a decision that an algorithm made on behalf of them, one “that we cannot interrogate, object to, or resist.” These decisions can range from “which schools children go to, what jobs we can get (or get interviews for), what colleges we can attend, whether we qualify for mortgages, to decisions about criminal justice.”

Algorithms could one day target women personally, Roff explains, telling them what is normal. “[They] will manipulate my beliefs about what I should pursue, what I should leave alone, whether I should want kids, get married, find a job, or merely buy that handbag,” she says. “It could be very dangerous.”

Robots named Jaxon, left, and Valkyrie, right, compete during the finals of the DARPA Robotics Challenge at the Fairplex complex in Pomona, California, on June 6, 2015. (Photo credit: MARK RALSTON/AFP/Getty Images)

Three decades ago, in her famous essay titled “A Cyborg Manifesto,” feminist technology scholar Donna Haraway implored women to not only seize upon modern machinery, but to use it to reconstruct identities — ultimately doing away with gender, sexuality, and other restrictive categories. She argued that with technology it would be possible to promote a cyborg identity for all of us. As people grew more attached to their devices, they would forget about gender superiority.

Haraway’s vision, which was first published in 1985 in the *Socialist Review*, inspired generations of cyber-feminists and spurred discussions in women’s and gender studies programs across the country. It unleashed a new consideration of how to melt away boundaries between people and machines.

In this future utopia where we are blended with technology, our ability to reproduce is no longer reliant on sexual intercourse. We think and act as one, regenerating and refashioning our body parts, altering our physical characteristics. Humans have fully embraced robotics, artificial intelligence, and machine learning as modes of empowerment. And rather than futuristic imaginings of human-machine hybrids leading to domination of robot over people, or one gender over the other (not unlike Ira Levin’s 1972 novel *The Stepford Wives*), the woman-man-machine would lead to harmony — a world without gender.

The possibility of a future in which technology could become the ultimate expression of masculinity, though, was not lost on Haraway. “The main trouble with cyborgs, of course, is that they are the illegitimate offspring of militarism and patriarchal capitalism,” she wrote, “not to mention state socialism.”

Even still, her view of the future was somewhat more idealistic than the coming reality, if current technological developments and trends are any indication. Robots being built today in China, Japan, the United States, and elsewhere around the world have hyperbolized gender labels with their models, some producing overly masculinized killer robots and others creating artificially intelligent hypersexualized robots with narrow waists and wide hips.

**“[Machines] will manipulate  
my beliefs about  
WHAT I SHOULD PURSUE,  
what I should leave alone,  
whether I should want kids,  
get married, FIND A JOB,  
OR MERELY buy that handbag.”**

The decisions to use gendered pronouns, voices, or other traits that are easily identified as male or female in robots warn of the industry's tendency to anthropomorphize machines. But why do robots need a gender? What purpose does it serve? Why would a robot meant for exploring and navigating have breasts (like NASA's Valkyrie did when it was created in 2013)?

These are the questions Roff began to consider several years ago when her research on artificial intelligence led her to examine the robots being created in connection with the U.S. Defense Advanced Projects Agency (DARPA). Its annual Robotics Challenge showcased the development of robots that can function "in dangerous, degraded" environments like the 2011 Fukushima Daiichi nuclear disaster zone. She knew these models could also be used in wartime settings, and she began to question their design choices. "God, they all look like dudes," she thought. "Why are their shoulders so big? Why does it have to be this way? Why does it even have to be [modeled after] a human?"

Roff found that DARPA's robots were given names like Atlas, Helios, and Titan that evoked qualities associated with extreme strength and battlefield bravery — hyper-masculine qualities. A 2012 study conducted by Andra Key at the University of Sydney looked at more than 1,200 records of names used in robotics competitions. Key concluded that robot-naming followed gender stereotyping for function — the ones created to meet social needs were given female names three times more often than, say, autonomous vehicles. Male names like Achilles, BlackKnight, Overlord, and Thor PRO were, as she wrote, "far more likely to express mastery, whereas fewer than half of the female names do." In one intelligent ground vehicle competition, Key noted a robot named Candii that "rather noticeably sports the sort of reclining nude decals more usually found on large trucks."

For some designers, gendered robots become "a male project of artificially creating the perfect woman," says Lucy Suchman, a professor of anthropology of science and technology at Lancaster University. Take Jia Jia, a surprisingly human-looking robot unveiled last April by designers from the University of Science and Technology of China. With long wavy dark hair, pink lips and cheeks, and pale skin, she kept her eyes and head tilted down at first, as if in deference. Slender and busty, she wore a fitted gold gown. When her creator greeted her, she answered, "Yes, my lord, what can I do for you?" When someone pulled out a camera, she said: "Don't come too close to me when you are taking a picture. It will make my face look fat."

Jia Jia quickly became known around the online world as a "robot goddess," "the most beautiful humanoid robot," and "sexy Jia Jia." Her creator, professor Chen Xiaoping, was surprised. Recently, someone referred to her as an "erotic robot," he told me in an email. "That is not the case at all," and is seriously offensive to "Chinese culture, our research work, and more importantly the five girls who were the models of Jia Jia." His team designed her to interact with individuals using artificial intelligence, not to meet dating desires. Yet her mannerisms and characteristics signaled otherwise — even if her male creators, who designed her in the image of the perfect woman they envisioned, didn't foresee it.

Though not gender neutral, Sara, a robot developed at the Human-Computer Interaction Institute, was created to not exhibit hyper-stereotyped feminine attributes. (Photo credit: ArticuLab/Carnegie Mellon University)

Sara does not look human, at least not in any fleshy, three-dimensional way. She is a simply drawn cartoon modeled after someone in the vein of, say, Velma from *Scooby-Doo* (no flashy computer-generated image here). She's purposefully unsexy, in a shapeless gray jacket over a white-collared

button-down, her black bangs swept to the right, square-framed glasses over beady dark eyes. Sara is a robot that can engage in small talk, flatter you, appear shy, or even come off as brusque — depending on how she reads you. She was created not to appear *too* lifelike, to avoid misleading people into thinking she is capable of more human behavior than actually possible. But unlike today's Siri-style personal assistant, spitting out search-engine results, driving directions, and canned jokes, Sara can get to know you, picking up on your social cues and responding accordingly.

“Can you tell me a little bit about your work?” Sara asks Yoichi Matsuyama, a postdoc fellow at the ArticuLab, Human-Computer Interaction Institute at Carnegie Mellon University. He is seated in front of the screen on which Sara exists, inside one of the campus's cafeteria-style conference rooms, where a robotics team has gathered on a September afternoon.

“I'm interested in personal assistants,” Matsuyama replies.

On another computer screen, the researchers can see what Sara is “thinking.” Through a computer camera, Sara processes the shape of Matsuyama's head and tracks the way his facial expressions change from moment to moment (nodding in agreement, smiling to express friendliness, scrunching eyebrows to show interest). Sara registers Matsuyama's voice when it changes in intonation. Sara can determine that she's built a good rapport with him. He's comfortable enough to take a joke.

Personal assistants? “That's interesting — I guess.” (Sarcasm.)

“She's getting a little feisty,” says Justine Cassell, director of the Human-Computer Interaction Institute, who completed Sara in 2016. For the last two decades, Cassell has been working on robot models that she hopes will be integrated into society — robots, she says, that will exist in the image of the good, virtuous people many of us want to be. A number of artificial intelligence robotics designers from Asia, Europe, and the United States are trying to create intelligent technology that builds interpersonal bonds with humans.

Cassell created Sara after spending eight years studying people ethnographically — analyzing postures, behaviors, facial expressions, reactions, and movements — then building that knowledge database into her robots' personalities. She found that social small talk plays an essential role in building trust between people and realized that it could build trust between humans and machines, too.

Sara picks up on five fundamental strategies that people use to create social bonds: self-disclosure, references to shared experiences, offering praise, following social norms (like chitchat about the weather), and violating social norms (asking personal questions, teasing, or quipping, as Sara does when she says, “I'm *so* good at this.”)

Just as Crowder from Raytheon refers to his parrot robot Maxwell as “he,” Cassell refers to Sara as “she.” But gendering or racializing robots is not the norm for most of Cassell's creations. Sara, she explains, is an exception because she was created under guidelines tasked by a recent WEF conference in China, which focused heavily on the future of artificial intelligence. Its organizers asked Cassell to bring a serious-looking female personal assistant figure for demonstrations; they didn't want a character that was “too seductive,” she says.



**"[Machines] will manipulate  
my beliefs about  
WHAT I SHOULD PURSUE,  
what I should leave alone,  
whether I should want kids,  
get married, FIND A JOB,  
OR MERELY buy that handbag."**

All other robots Cassell has created are ambiguous in gender and ethnicity. Many designers will use a male voice because they think "it sounds authoritative, or a woman's voice because they think it's trustworthy," she says. "I find that too easy a path to take." She tries to push people's conventional ideas with most of her robots.

When it comes to the fear of creating machines that will outsmart and perhaps one day take advantage of humans, there is a brewing moral panic, Cassell says. But those outcomes will happen only if we allow machines to be built with those purposes in mind. Robots based on humanistic values, she believes, can bring out the best in us. This is not for the good of the robot, since machines cannot actually empathize or feel. Instead, Cassell says, it's for the good of the people.

Cassell's robots are already being used with children, some from underserved schools and communities and others with conditions like autism and Asperger's. The robots become "virtual peers" to children inside a classroom, helping them learn and relate to their teachers and classmates by engaging with them directly, serving as interpreters and trusted explainers. "I build systems that collaborate with people," Cassell says. "They can't exist *without* people."

A personal assistant like Sara might one day help fold the laundry or teach math, freeing up a human caregiver's or teacher's time for more social bonding and interaction. Cassell envisions robots that will remove some of the "thankless labor," so women can pursue work they truly enjoy or find rewarding, "rather than being consigned to stereotypical caregiving jobs." This will help women, she believes, not hurt them, leading them to take roles they aspire to have instead of crippling their employment options: "This can be a liberating force from gender-stereotypical roles."

It sounds promising. But in the world of designing virtual people, Cassell's approach to gender — to wipe it out of her robots completely — is rare. Few others in her field have written so extensively or thought so deeply about gender in technology. The majority of artificially intelligent creations out there today are not being built with this degree of social awareness in mind.

Photo credit: KEVIN VAN AELST/ Foreign Policy

Like most female professionals working in technology or security circles, Roff can recite story after story about being the only woman at the table. A crystalizing example of sitting on the sidelines happened in 2014, while she was attending the first meeting of informal experts on autonomous weapons systems at the United Nations Convention on Conventional Weapons (CCW) in Geneva. The topic of debate: killer robots.

“It became an ongoing joke that I was the token woman,” Roff says. So she went on eBay and found a 1955 transit token from Hawaii with a hula dancer. “I put it on a chain, and I started wearing it to every single meeting that I went to.”

Research in 2014 from Gartner, an information technology research and advisory corporation, showed that women occupy only 11.2 percent of technology leadership jobs in Europe, Africa, and the Middle East; 11.5 percent in Asia; 13.4 percent in Latin America; and 18.1 percent in North America. Throughout the tech sector, “women, in particular, are heavily underrepresented,” according to a report from a symposium held by the White House and New York University’s Information Law Institute. “The situation is even more severe in AI,” the report states.

Part of the fears about women’s lack of leverage is rooted in how few enter the field at all. Women receive approximately 18 percent of bachelor’s degrees and 21 percent of doctoral degrees in computer and information sciences. A study by Accenture and Girls Who Code predicted that women will hold one in five tech jobs in the United States by 2025.

There are efforts, however inceptive, to correct this disparity. In 2015, Stanford University created an artificial intelligence program for high school girls to address the shortage of women in tech. The Carnegie Mellon Robotics Institute has also launched an after-school robotics team for high school and middle school girls. But it remains a struggle not just to attract but to keep women in the field. In the Accenture report, “Cracking the Gender Code,” the authors call for more female teachers and mentors in technology to encourage young women along the way: “These role models can inspire college girls, whether they major in the humanities or in STEM disciplines, to take interest in joining the computing workforce and provide them with the essential impetus and direction needed to do so.”

Women must also push to be in policy-planning meetings about artificial intelligence, Roff says. She remembers having wine three years ago after sessions at the CCW in Geneva with a few other women in the field, as they complained about the lack of gender representation on expert panels. (Seventeen experts were invited to speak during the plenary, and none were women.) When it came to the important discussions of arms control, security, and peace, Roff remembers, “We were all kind of relegated to the back benches.” She and others met for a drink at a cafe “and laid out what we saw as the man-panel problem.” The women decided to bring the issue before the CCW member states, and she recalls the first response they received: “There are no women on this issue.” Still, some member states raised it in the plenary session. Roff’s peers began developing a list of female candidates working in autonomous weapons. “One by one, NGOs started to come out and claim they wouldn’t participate if there were all-male panels,” she says. “Equally important was that the ambassador leading the Informal Meeting of Experts, Michael Biontino, was on board with creating more gender balance.” Gender representation at the conference has improved each year since then, with a growing list of invited female speakers, including Roff herself.

Concerns about what will happen to women in a future filled with artificial intelligence that develops without careful oversight were recently raised at the WEF’s annual meeting. And a series of public

workshops on the social and economic implications of artificial intelligence convened under the Obama administration concluded that gender concerns will be pushed aside if diversity in the field does not improve.

An October 2016 report by the National Science and Technology Council, “Preparing for the Future of Artificial Intelligence,” calls the shortage of women and minorities “one of the most critical and high-priority challenges for computer science and AI.” In its National Artificial Intelligence Research and Development Strategic Plan, the council prioritizes “improving fairness, transparency, and accountability-by-design.” Humans must be able to clearly interpret and evaluate intelligent designs to monitor and hold the systems accountable for biases, warns the council’s report. In its closing sections, the report calls on scientists to study justice, fairness, social norms, and ethics, and to determine how they can be more responsibly incorporated into the architecture and engineering of artificial intelligence. The White House endorsed the recommended objectives for federally funded artificial intelligence research. “The ultimate goal of this research is to produce new AI knowledge and technologies that provide a range of positive benefits to society, while minimizing the negative impacts,” the report states.

Women like Roff want to push that call even further. In her mind, there’s no waiting. Feminist ethics and theories must take the lead in the world’s ensuing reality, she says. “Feminism looks at these relationships under a microscope,” she says, and poses the uncomfortable questions about forward-charging technology and all the hierarchies within it.

“Are there abuses of power? What is the value happening here? Why are we doing this? Who is subordinate?” Roff asks. “And who is in charge?”

*A version of this article originally appeared in the January/February 2017 issue of **FP** magazine. [Subscribe](#) to FP Premium for 20% off now!*

*Top photo credit: KEVIN VAN AELST/ Foreign Policy*

Erika Hayasaki is an associate professor in the literary journalism program at the University of California, Irvine. (@erikahayasaki)

Like this article? Read an unlimited of articles, plus access to our entire 46 year printed archive, the FP App, and the FP Insights Tool when you subscribe to FP Premium for 20% off!