

‘I do not think ethical surveillance can exist’: Rumman Chowdhury on accountability in AI

Publication Date: 2023-05-29

Author: Unknown

Section: Technology

Tags: Artificial intelligence (AI), US politics, interviews

Article URL: <https://www.theguardian.com/technology/2023/may/29/rumman-chowdhury-interview-artificial-intelligence-accountability>



Rumman Chowdhury often has trouble sleeping, but, to her, this is not a problem that requires solving. She has what she calls “2am brain”, a different sort of brain from her day-to-day brain, and the one she relies on for especially urgent or difficult problems. Ideas, even small-scale ones, require care and attention, she says, along with a kind of alchemic intuition. “It’s just like baking,” she says. “You can’t force it, you can’t turn the temperature up, you can’t make it go faster. It will take however long it takes. And when it’s done baking, it will present itself.” It was Chowdhury’s 2am brain that first coined the phrase “moral outsourcing” for a concept that now, as one of the leading thinkers on artificial intelligence, has become a key point in how she considers accountability and governance when it comes to the potentially revolutionary impact of AI. Moral outsourcing, she says, applies the logic of sentience and choice to AI, allowing technologists to effectively reallocate responsibility for the products they build onto the products themselves – technical advancement becomes predestined growth, and bias becomes intractable. “You would never say ‘my racist toaster’ or ‘my sexist laptop’,” she said in a Ted Talk from 2018. “And yet we use these modifiers in our language about artificial intelligence. And in doing so we’re not taking responsibility for the products that we build.” Writing ourselves out of the equation produces systematic ambivalence on par with what the philosopher Hannah Arendt called the “banality of evil” – the wilful and cooperative ignorance that enabled the Holocaust. “It wasn’t just about electing someone into power that had the intent of killing so many people,” she says. “But it’s that entire nations of people also took jobs and positions and did these horrible things.” Chowdhury does not really have one title, she has dozens, among them Responsible AI fellow at Harvard, AI global policy consultant and former head of Twitter’s Meta team (Machine Learning Ethics, Transparency and Accountability). AI has been giving her 2am brain for some time. Back in 2018 Forbes named her one of the five people “building our AI future”. A data scientist by trade, she has always worked in a slightly undefinable, messy realm, traversing the realms of social science, law, philosophy and technology, as she consults with companies and lawmakers in shaping policy and best practices. Around AI, her approach to regulation is unique in its staunch middle-ness – both welcoming of progress and firm in the assertion that “mechanisms of accountability” should exist. Effervescent, patient and soft-spoken, Chowdhury listens with disarming care. She has always found people much more interesting than what they build or do. Before skepticism around tech became reflexive, Chowdhury had fears too – not of the technology itself, but of the corporations that developed and sold it. As the global lead at the responsible AI firm Accenture, she led the team that designed a fairness evaluation tool that pre-empted and corrected algorithmic bias. She went on to start Parity, an ethical AI consulting platform that seeks to bridge “different communities of expertise”. At Twitter – before it

became one of the first teams disbanded under Elon Musk – she hosted the company’s first-ever algorithmic bias bounty, inviting outside programmers and data scientists to evaluate the site’s code for potential biases. The exercise revealed a number of problems, including that the site’s photo-cropping software seemed to overwhelmingly prefer faces that were young, feminine and white. This is a strategy known as red-teaming, in which programmers and hackers from outside an organization are encouraged to try and curtail certain safeguards to push a technology to “do bad things to identify what bad things it’s capable of”, says Chowdhury. These kinds of external checks and balances are rarely implemented in the world of tech because of technologists’ fear of “people touching their baby”. She is currently working on another red-teaming event for Def Con – a convention hosted by the hacker organization AI Village. This time, hundreds of hackers are gathering to test ChatGPT, with the collaboration of its founder OpenAI, along with Microsoft, Google and the Biden administration. The “hackathon” is scheduled to run for over 20 hours, providing them with a dataset that is “totally unprecedented”, says Chowdhury, who is organizing the event with Sven Cattell, founder of AI Village and Austin Carson, president of the responsible AI non-profit SeedAI. In Chowdhury’s view, it’s only through this kind of collectivism that proper regulation – and regulation enforcement – can occur. In addition to third-party auditing, she also serves on multiple boards across Europe and the US helping to shape AI policy. She is wary, she tells me, of the instinct to over-regulate, which could lead models to overcorrect and not address ingrained issues. When asked about gay marriage, for example, ChatGPT and other generative AI tools “totally clam up”, trying to make up for the amount of people who have pushed the models to say negative things. But it’s not easy, she adds, to define what is toxic and what is hateful. “It’s a journey that will never end,” she tells me, smiling. “But I’m fine with that.” Early on, when she first started working in tech, she realized that “technologists don’t always understand people, and people don’t always understand technology”, and sought to bridge that gap. In its broadest interpretation, she tells me, her work deals with understanding humans through data. “At the core of technology is this idea that, like, humanity is flawed and that technology can save us,” she says, noting language like “body hacks” that implies a kind of optimization unique to this particular age of technology. There is an aspect of it that kind of wishes we were “divorced from humanity”. Chowdhury has always been drawn to humans, their messiness and cloudiness and unpredictability. As an undergrad at MIT, she studied political science, and, later, after a disillusioning few months in non-profits in which she “knew we could use models and data more effectively, but nobody was”, she went to Columbia for a master’s degree in quantitative methods. In the last month, she has spent a week in Spain helping to carry out the launch of the Digital Services Act, another in San Francisco for a cybersecurity conference, another in Boston for her fellowship, and a few days in New York for another round of Def Con press. After a brief while in Houston, where she’s based, she has upcoming talks in Vienna and Pittsburgh on AI nuclear misinformation and Duolingo, respectively. At its core, what she prescribes is a relatively simple dictum: listen, communicate, collaborate. And yet, even as Sam Altman, the founder and CEO of OpenAI, testifies before Congress that he’s committed to preventing AI harms, she still sees familiar tactics at play. When an industry experiences heightened scrutiny, barring off prohibitive regulation often means taking control of a narrative – ie calling for regulation, while simultaneously spending millions in lobbying to prevent the passing of regulatory laws. The problem, she says, is a lack of accountability. Internal risk analysis is often distorted within a company because risk management doesn’t often employ morals. “There is simply risk and then your willingness to take that risk,” she tells me. When the risk of failure or reputational harm becomes too great, it moves to an arena where the rules are bent in a particular direction. In other words: “Let’s play a game where I can win because I have all of the money.” But people, unlike machines, have indefinite priorities and motivations. “There are very few fundamentally good or bad actors in the world,” she says. “People just operate on incentive structures.” Which in turn means that the only way to drive change is to make use of those structures, ebbing them away from any one power source. Certain issues can only be tackled at scale, with cooperation and compromise from many different vectors of power, and AI is one of them. Though, she readily attests that there are limits. Points where compromise is not an option. The rise of surveillance capitalism, she says, is hugely concerning to her. It is a use of technology that, at its core, is unequivocally racist and therefore should not be entertained. “We cannot put lipstick on a pig,” she said at a recent talk on the future of AI at the New York University Silver School of Social Work. “I do not think ethical surveillance can exist.” Chowdhury recently wrote an op-ed for Wired in which she detailed her vision for a global governance board. Whether it be surveillance capitalism or job disruption or nuclear misinformation, only an external board of people can be trusted to govern the technology – one made up of people like her, not tied to any one institution, and one that is globally representative. On Twitter, a few users called her framework idealistic, referring to it as “blue sky thinking” or “not viable”. It’s funny, she tells me, given that these people are “literally trying to build sentient machines”. She’s familiar with the dissonance. “It makes sense,” she says. We’re drawn to hero narratives, the assumption that one person is and should be in charge at any given time. Even as she organizes the Def Con event, she tells me, people find it difficult to understand that there is a team of people working together every step of the way. “We’re getting all this media attention,” she says, “and everybody is kind of like, ‘Who’s in charge?’ And then we all kind of look at each other and we’re like, ‘Um. Everyone?’” • This article was amended on 1 June 2023 because an earlier version misnamed the New York University Silver School of Social Work, as New York University’s School of Social Sciences.