

Me, Myself, and AI Podcast

## Protecting Society from AI Harms: Amnesty International's Matt Mahmoudi and Damini Satija (Part 2)

**SAM RANSBOTHAM:** What can corporations learn from an activist organization that works to protect people from the harms of AI? Find out on today's episode.

**DAMINI SATIJA:** I'm Damini Satija ...

**MATT MAHMOUDI:** ... and I'm Matt Mahmoudi from Amnesty International ...

**DAMINI SATIJA:** ... and you're listening to *Me, Myself, and AI*.

**SAM RANSBOTHAM:** Welcome to *Me, Myself, and AI*, a podcast on artificial intelligence in business. Each episode, we introduce you to someone innovating with AI. I'm Sam Ransbotham, professor of analytics at Boston College. I'm also the AI and business strategy guest editor at *MIT Sloan Management Review*.

**SHERVIN KHODABANDEH:** And I'm Shervin Khodabandeh, senior partner with BCG and one of the leaders of our AI business. Together, *MIT SMR* and BCG have been researching and publishing on AI since 2017, interviewing hundreds of practitioners and surveying thousands of companies on what it takes to build and to deploy and scale AI capabilities and really transform the way organizations operate.

**SAM RANSBOTHAM:** Welcome back, everyone. On our last episode, Damini and Matt joined us to share a bit about what their organization, Amnesty Tech, is doing to combat troublesome uses of AI. Today, we're picking up on that discussion and sharing more detail about how you can be more aware of the dangers of artificial intelligence and, importantly, how you can help.

Damini, let's pick up where we left off last episode. For our listeners, we recommend you go back first and listen to our last episode, if you haven't yet, to get some more context on Amnesty Tech and the work that Matt and Damini are doing.

Damini, you and Matt were starting to talking about AI regulations and how they can help us address challenging tech problems, like housing algorithms, social work, and facial recognition systems. Let's pick up from there. Can you share more about your perspective on regulating AI?

**DAMINI SATIJA:** Regulation is a key part of the toolkit here. We're working really hard on the EU's Artificial Intelligence Act, which is right now one of the most comprehensive frameworks out there for regulating AI. I think what's really important with regulation, and what we're really missing

right now, are regulatory frameworks, which really focus on the outcomes that we want

to prevent or even promote. And by that, what I mean is that a lot of the regulation we're seeing — even in the AI Act, which is a very advanced piece of AI regulation — but even in the case of the AI Act, it's very often tied to sort of tech hype cycles and the technology that is the hype of the moment.

And the way we've seen this really clearly with the AI Act is that in the last few weeks and months, as the conversation has really picked up around generative AI, we've seen policy makers, who are deep in the AI Act, which is really in its last phases, not know how to absorb generative AI into the framework. And I think we don't have a very robust regulatory framework if it cannot absorb a new technological development. And that's not what the goal was. In the early days, there was a lot of work done upfront with the AI Act, saying we want to "future-proof" this regulation. It will be an instrument that is ready to impose the restrictions and protections we need as the technology develops. But right now, it seems like it's not doing that.

And I think that's because the regulation attempt itself is so tied to the technology hype cycle, as I say, and what we need is to be more focused on the outcomes we want to prevent, and so many of those outcomes are embedded in the way we think about human rights — so the right to nondiscrimination, the right to privacy. There are certain outcomes we know we need to get to to protect human rights,

regardless of what the technology we're talking about is. So that's what I would add on the regulation front and what I think is really missing right now.

I'd also add to the urgency for this, given the rapid pace of technological development, but also, slightly tangentially, algorithmic and AI and technology in general [being] picked up in public-sector environments — which is much of my focus [and] a lot of Matt's focus — in these restrained environments. They've been called "austerity machines" for that reason. And given where the world is right now in the latest stages of the pandemic [and] the global economy seeing multiple shocks, we can very easily anticipate that these austerity machines could become even more commonplace. And that's why this applies to AI in general, but just thinking about the area that my team works very specifically on in the welfare and social protection context, that urgency feels very dire right now.

And secondly, these efficiency tools are often designed to detect or weed out fraudulent applicants for welfare and public services, so these are really punitive tools as well, in the name of efficiency. That's where the disproportionate impact happens on low-income groups, communities of color, etc. So this entire narrative drives really harmful outcomes. We see that narrative only accelerating, given the context that we're in. And so the case and the urgency for that regulation is very strong right now.

**SAM RANSBOTHAM:** I think that's pretty interesting. Part of my backstory involves ...

I used to work for the International Atomic Energy Agency in Vienna, and that's a nuclear regulatory [organization]. You can point to a lot of difficulties with that model, but we have not had nuclear explosions since 1945, or, you know, we've not had nuclear warfare.

But Matt, you also mentioned local regulation. And this was an idea — and even the EU is, let's say, country- or "group of countries"-related — but it seems like this is going to have to be something at the global level versus that local level. Or, where do you see the level of this regulation taking place? With nuclear, it seemed to require [it] at the global level, and it did a nice job, I think — I'm biased — but it did a nice job of pairing positive uses of the technology with limiting the negative uses. What level should this regulation be, then?

**MATT MAHMOUDI:** I mean, there's no counterargument to say that there shouldn't be global legislation on this or there shouldn't be global-level agreements and resolutions in place on this that impose binding obligations on states when it comes to the deployment and development and the usage of AI — not just in civilian matters, but in the context of warfare as well.

However, I think as far as the most progressive and immediate-term impacts we've seen when it's come to advocacy in terms of regulation, it has been at the local level because constituents are very good at activating their local lawmakers toward taking decisive action at the, for example,

city council level. We're seeing movements at the New York City Council level, as we speak, toward moving for a ban on the usage of facial recognition in residential housing.

We've also seen movements that will be introduced later on that we'll be speaking about in the context of law enforcement. We've seen, in the context of Portland, Oregon, legislation being put into place. At a moment that was so critical ... especially leading up to the Black Lives Matter movement, with the murder of George Floyd, where the kinds of racial impacts and racializing impacts of these technologies were becoming even more clear, that by allowing the deployment of facial recognition, you're not simply allowing the usage of an experimental tool that has more of a tainted record than it has a record showing sort of positive affordances of any sort, but you're also enabling the existing sort of institutionalized racism that does exist within police forces to be put on steroids, in a great many ways. And, of course, a lot of the claims that protestors were making during this moment were against police abuse. And so you can't have a challenge to policing and then also facial recognition. So that's all to say that the local level will drive a lot of the demand, even at the global level, for regulation.

And I think in stitching those pieces together and being able to draw these stories that — look, there is no instance of X,Y, Z form of technology or AI-driven surveillance in any context that has shown us that we can safely just take our hands off

the wheel and let it do its thing. That is what's going to galvanize the kinds of regulation we might want to see at that level. And the kinds of "EU effect" that certain civil society organizations refer to might be something to look for: the ways in which regulation and regulatory models jump from one space to another.

I will also say that there have been processes at the UN that speak specifically to the usage of autonomous weapons systems, which has been a long-winded process so far but which does seek to address issues of AI in the context of warfare.

**SHERVIN KHODABANDEH:** Most of our podcasts to date have been on how these tools could help [create] dramatic improvements in efficiency and effectiveness and also do positive for the world and for the environment. There's another side of this, which Damini and Matt have so eloquently shone a light on, [and] that is the power imbalance of the usage and the outcomes. And I think it's an important dialectic that has to happen over time.

So as much as I'd like to push more on "Can technology at least be part of the solution?" — [and] I fully believe technology is part of the solution — I think the existential nature of the issue is such that you need to have this dialogue and this discourse.

Matt, you talked about image recognition, right? If you look at image recognition improvement over time, it has improved exponentially. To date, it still creates problems when it's used at such a wide scale, right? Obviously, if it's got [a more] close-up view, it might not. But imagine a world, maybe 10 years from now, maybe 20, who knows, where the instrumentation is so far advanced, and the algorithms are so far advanced, and the safeguards are there, that it actually trumps a human — you know, the very people that went and counted [New York City police surveillance] cameras [visible in Google Street View images] — trumps their ability to tell the difference between people.

I would assume in that world, you'd be OK with it being used ... or not? What would be your view — let's say, if you were to project 10 years from now — if some of these tools just don't make mistakes anymore? And now you only have the bad-actor situation, but the tool itself does not make a mistake. Would that change your position?

**MATT MAHMOUDI:** That's terrifying. And to me that's terrifying because it ...

**SHERVIN KHODABANDEH:** It is terrifying.

**MATT MAHMOUDI:** It is! It creates the conditions under which institutions, which are imperfect and [represent] varying positions on an ethical spectrum, are suddenly in power to do things at great scale, with great precision. So it's no longer that you have the NYPD being able to just find whoever they can using facial

recognition that was provided to them as sort of a test case. It's that they can go and target someone specifically. And they're able to do so at a massive scale, no longer with the kinds of false positives that we've heard about from Detroit and New Jersey and elsewhere, but, like, actually enabling them to carry out to fruition the existing ...

And there's data on this, right? There is data to support that, for example, stop-and-frisk incidents target upwards of 90% Black and Brown people and that these happen in mostly Black and Brown neighborhoods. That is not because Black and Brown neighborhoods are predominantly full of crime. That is because that is where the targeting happens, and so there is greater visibility. And as it so happens, most of the cases of the stop-and-frisk incidents don't actually lead to an arrest. So that, again, shows you that there is no sort of credence to the idea that these communities are inherently criminal, in any [type] of way.

So then imagine a future in which the police are empowered to do exactly that — that is to say, a digital stop and frisk. Everyone is virtually lined up, without their knowledge and consent, simply because that's how this institution operates. And now it's given free rein to do so at a scale that it hasn't been able to do before. That is terrifying to me. We can't have that.

I think that's a very real scenario that we have to consider, and that has profound implications for Americans' First Amendment rights, for the rest of the

planet's right to protest. And that becomes harder. And so, what do we do when we're faced with, say, for example, a state or a government that has suddenly fallen out of favor with, say, its populace, but it has been equipped with these of awe-inspiring technologies of horror? How do you then challenge that government, if there aren't protections in place to ensure that those technologies weren't given to them in the first place, for example? That would be my contention as to why this is a terrifying scenario.

**SHERVIN KHODABANDEH:** Yeah, I think you're right. I think you're right. And this is why I actually ... I think this was very helpful, because you cannot de-link the technology from the user and the imbalance of power and the fundamental possibility of corruption in certain situations.

**SAM RANSBOTHAM:** You mentioned the state and the government entities. You haven't even brought in the corporate world, where so much of this is happening in a very concentrated hegemony of power that, whatever oversight you may be concerned about with the NYPD's and some of your examples, I think we have that on steroids with the lack of oversight or ability to control what happens within the darkness of a corporation. And then that's true even with well-intended people there. That's not ascribing necessary malice. It's just ascribing self-interest.

And as I think about the power imbalance, we as individual people have so very little

with any of these other collectives. There's exactly zero of these artificial intelligence algorithms, the recommendation engines, that care about what I want to see. They care about what the corporation or the advertiser or whoever. Their objective function is fundamentally not my objective function. Now, often it's similar enough to where it's OK. But inherently, their objective function is not necessarily my objective function, and that's where the power imbalance seems really difficult.

**SHERVIN KHODABANDEH:** Damini, I'm going to also ask you, 10 years from now, how do you see the future here? What would be good? What would be terrifying? Matt depicted a very terrifying future.

**DAMINI SATIJA:** Yeah, I would hope for the nonterrifying future. I think the question, to me, gets back to this question of power. And we've all mentioned it now, and the power imbalances in the current technological trajectories that we see. Yes, corporate power and government power — and those two things are not disconnected either, right? Like, where do governments get these technologies? [They're] actually increasingly less and less developed in-house and [more often] procured from somewhere. Government creates demand for these technologies, companies solve

these ... It's all connected, and it's all part of the system of, where does the power sit? From start to finish, who is investing in what technology? How is that technology being developed? Who decides where it's deployed, where it's sold, who's buying it?

If I try and envision the future 10 years from now, which is not the terrifying future that you and Matt have discussed, it is the opposite of that. It's one where we're able to dismantle some of those power structures which drive the current trajectory of technological development. It's where we're able to give power to those who typically have not had voice in what technologies are developed and how they're used to their benefit. Because I do believe, to an earlier point, that there are ways we can use these technologies and develop these technologies to really lift people out of positions of systemic disadvantage and marginalization in society, but we need to bring out the visions for how that can happen. And right now there is no way for those visions to exist. There is no time for those of us working on sort of human rights impacts of tech — but, more importantly, those impacted by the use of these technologies — to put those visions out there.

So I don't have a specific future to give you or outline for you but rather a way that I would like to move toward defining what that future looks like. And I think in order to do that, it's really important for us to not always take a position of "What are the benefits and what are the risks of each technology?" and force ourselves into a position of assessing every new technological development we see from a place of this balance. Because we've been doing that so far and it's led us to a point at which we have these corporations with huge hegemonic power.



Social media is a really good example of this. I think for years now, we forced ourselves to say, “Yes, there are all these bad things happening, or that we can foresee happening, but also think about all these benefits that social media has brought to society.” Yes, but there is a reason that we need to focus on those harms. There is a reason that we need to shine more light on those risks, because if we don’t, we’re not seeing where we need to shift the way that technology is developing and what red lines we need to draw and what we need to change.

**SAM RANSBOTHAM:** Most of our listeners are corporate or government workers. Right now, let’s say they buy into the things you’re saying. What should they be doing? What should they be thinking about? What should each person be doing right now?

**DAMINI SATIJA:** I mean, wherever you work, I think you have a responsibility — especially if you are building new technologies, you’re contributing to the development of tech, the use of tech. Matt and I work on some very specific contexts that we’ve talked about today, but as we’ve also mentioned in this call, there are so many domains in which tech is used that we don’t work on: education, health care, others. And these issues sit across all of those domains. So I would just say, think about your position, the power that you do have, the responsibility that you do have to bring these issues up internally, whatever organization or company you are at. I think people forget sometimes the power you can have in bringing these issues to light behind

closed doors. You know, some of our work is very public, but a lot of these conversations and the really important decisions are made behind closed doors.

**SAM RANSBOTHAM:** Matt, Damini, this has been a great and different discussion for us. It seems especially important as consumer AI tools start to proliferate. While we often focus on the positive ways organizations can use AI, positive uses, as we know now, are not the only ways that we can use AI. But the good thing is that we have considerable human agency in how we use these tools. Thanks for taking the time to talk with us.

**DAMINI SATIJA:** Thank you for having us on.

**SAM RANSBOTHAM:** That’s a wrap on *Me, Myself, and AI* Season 7. We’re blown away by how popular this show is, and we greatly appreciate all of our listeners. Please feel free to continue to make suggestions as we continue to grow. We’ll be back later this fall with more new episodes. In the meantime, please consider joining our LinkedIn community, and rate and review our show. Also, please suggest it to any friends or colleagues who might benefit from these conversations. We thank you for your support and will speak with you again soon.

**ALLISON RYDER:** Thanks for listening to *Me, Myself, and AI*. We believe, like you, that the conversation about AI implementation doesn’t start and stop with this podcast. That’s why we’ve created a group on LinkedIn specifically for listeners like you. It’s called AI for Leaders, and if you join us,



you can chat with show creators and hosts, ask your own questions, share your insights, and gain access to valuable resources about AI implementation from *MIT SMR* and BCG. You can access it by visiting [mitsmr.com/AIforLeaders](https://mitsmr.com/AIforLeaders). We'll put that link in the show notes, and we hope to see you there.