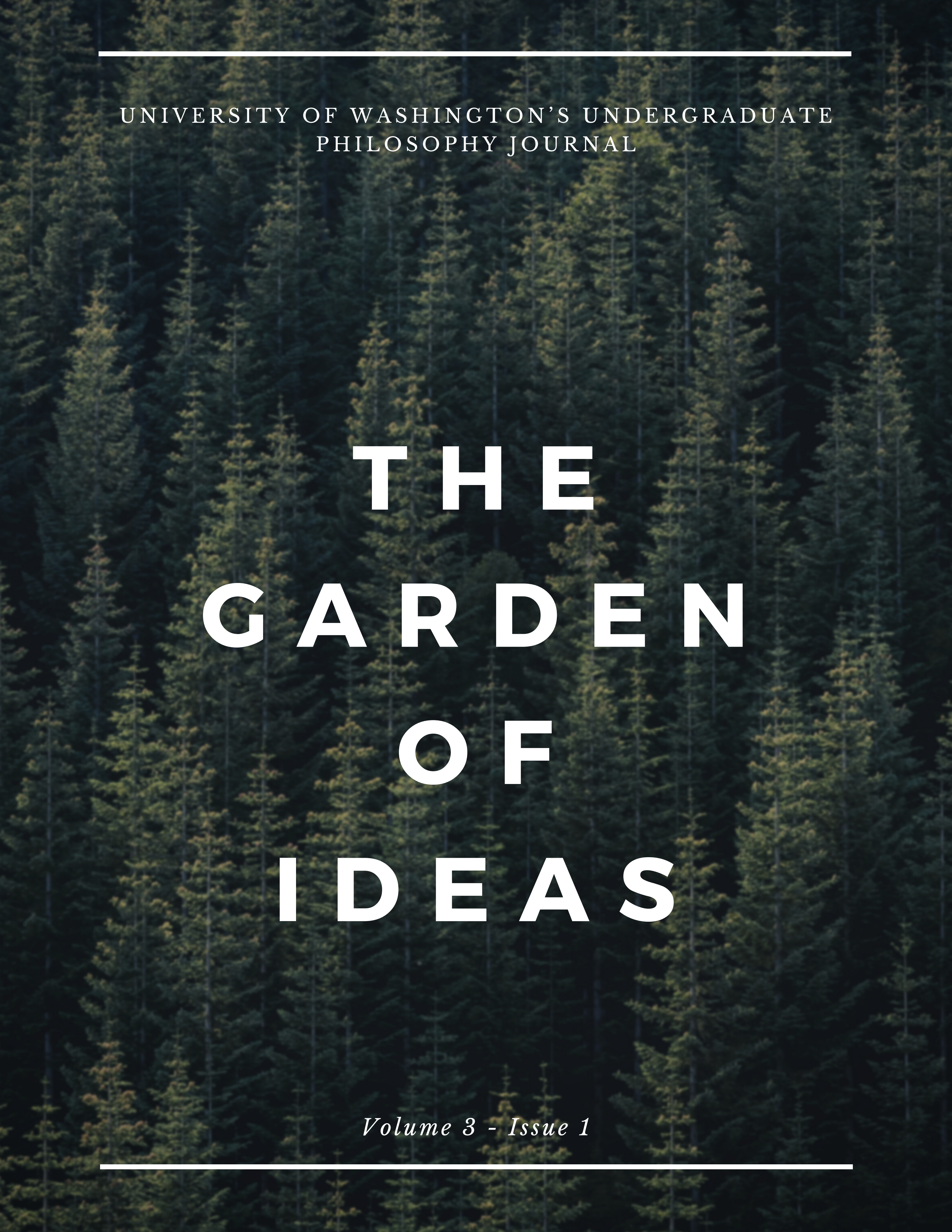
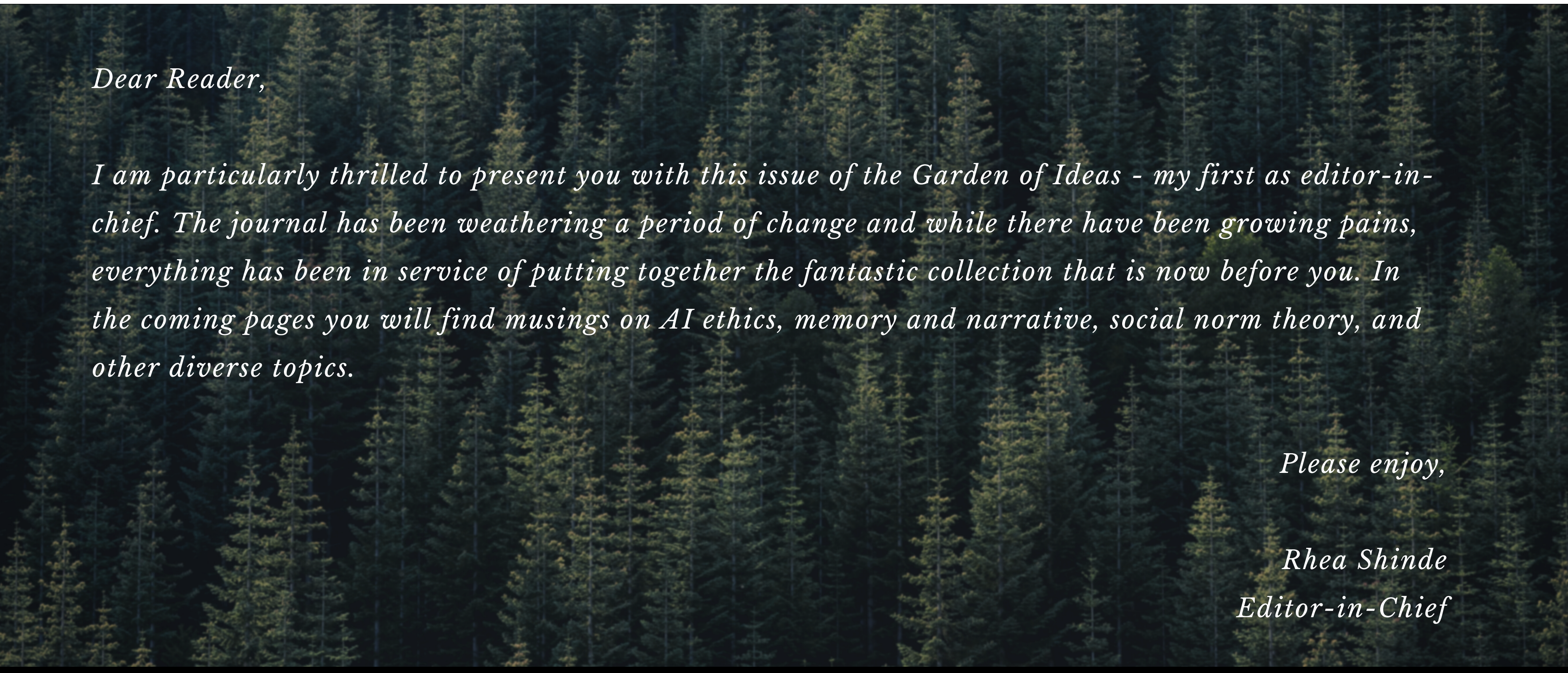


UNIVERSITY OF WASHINGTON'S UNDERGRADUATE
PHILOSOPHY JOURNAL

# THE GARDEN OF IDEAS

*Volume 3 - Issue 1*

# TABLE OF CONTENTS



*Dear Reader,*

*I am particularly thrilled to present you with this issue of the Garden of Ideas - my first as editor-in-chief. The journal has been weathering a period of change and while there have been growing pains, everything has been in service of putting together the fantastic collection that is now before you. In the coming pages you will find musings on AI ethics, memory and narrative, social norm theory, and other diverse topics.*

*Please enjoy,*

*Rhea Shinde*
*Editor-in-Chief*

# CARE ETHICS IN THE AGE OF AI

## AN INTERVIEW WITH DR. PETER ASARO

*Andrew Shaw and Molly Banks*

*The Garden of Ideas was very fortunate to be able to host a live interview with Dr. Peter Asaro on November 30, 2023, conducted by editors Andrew Shaw and Molly Banks. Dr. Asaro is a leading philosopher of science, technology, and media and currently a visiting scholar at the UW Center for an Informed Public. His research focuses on social, cultural, political, legal, and ethical dimensions of automation and autonomous technologies. The transcript below has been edited for print.*

**Molly: We had a lot of interest in your paper "AI Ethics and Predictive Policing: From Models of Threat to Ethics of Care," and in that paper you explored the theoretical and practical advantages of abandoning the more common models of threat approach in favor of the more holistic ethics of care approach. What are the limitations of a models of threat approach, specifically in the context of predictive policing?**

Dr. Asaro: Maybe it's also helpful to describe what the models of threat is, and what I was trying to do by characterizing a whole set of practices that I want to put under this umbrella "models of threat." This is a very utilitarian approach that you see in a lot of different domains. At the UN level, you have international relations, national security threats, international threats and analysis. Looking at policing, there's various kinds of threats. There's a longer history around threat modeling within cyber security, it goes back to the early 2000s. Within that, there's this engineering mentality/utilitarian mentality of "Identify the data and fix the problem. If the problem is too many errors, let's reduce the number of errors. If we don't have the right data, let's get more data."

This particular paper is from 2019, so there's a historic moment at which a lot of work was starting to come out about data and algorithmic bias, and ways in which learning systems or machine learning had all this intrinsic, implicit racial bias. So the reaction within the tech community is, let's just debias the data. We know that there's bias in it, we're just going to come up with some fancy computational techniques to fix it. I think it misses the point. You missed the forest for the trees by focusing on trying to make data more accurate, more precise. There was actually, at the FAccT (Fairness, Accountability, and Transparency) conference, this paper on 32 different definitions of fairness that might be applied. What counts as fair is debatable, philosophically. So it didn't seem like that was the right approach, but that seemed to be where all the conversation, all the resources were moving in AI ethics.

In this paper, I really want to critique that, take a step back, and think about, “how else can we think about AI ethics [in a way] that's not just utilitarianism, not just error reduction or debiasing?” I turned to feminist theory and the ethics of care, and then within the paper, I do a comparative analysis of two applications of data-driven policing that happened contemporaneously in the City of Chicago, where gun violence was extremely bad in the mid-decade. These two different programs both used data to try to identify who was at risk for gun violence, but then did radically different things with it. One is looking at it as threats. People who are likely to be involved in gun violence represent this threat, so we can police them more intensively. There's several very elaborate statistical analyses of the results, and they found it did not work at all: zero impact on gun violence for those people who were identified in that system. That system was the SSL (Statistical Suspect List), or hot list.

This other program identified at-risk youth by looking at high schools that were in the most gun violence-prone and the lowest socioeconomic status neighborhoods within the city, did a review of applications from those students, but then gave them summer jobs. Instead of what the police did with this SSL list, [where] they showed up at the house of people and threatened them. Or when there was an incident in the neighborhood, they would use that to generate a list of people with high ratings on this list and round them up for questioning just because they were proximate. But the one summer the City of Chicago gave them jobs with mostly community organizations, [it] basically reduced gun violence for the initial class by 51%—massive reduction for any kind of policy intervention. It didn't really depend on making the data more accurate. It wasn't, “if we collect a bunch more data, debias this data, or use fancier statistical methods, we're going to get better at predictive policing.” It's that they had a better plan for how to integrate that data into policies and, particularly, better policies about how to intervene to prevent violence. So this represents this idea of the ethics of care.

What you really have to think about, particularly in AI ethics, is how these systems are embedded within sociotechnical systems. It's not just the statistics that matter. It's the social structures in which they're embedded, and in this case, how police use it. How do you train police to understand how it works and what it does? They got no training and no instruction on what it really did. They're just told, “this will generate lists of likely suspects,” which actually wasn't true, because it lumped together people who might be victims with people who might be perpetrators, but they just treated everybody as potential perpetrators. But I think it applies more generally not just to policing but to many different things where we want to apply technology. We think, “Here's a threat. Here's a risk. How do we minimize that?” and then apply tons of computational power and data analysis to doing that, rather than thinking about, “what are the social implications, and how do we actually reframe the way people think about the world?” We're building this technology, which only has the capability of identifying threats, and that gives you only a set of actions to take that are a response to threat, which is aggression and arrest. Subsequently, with George Floyd, we've become much more aware that there are other ways to do policing, and police don't necessarily need to use violence to solve a problem. In many cases, people are psychologically disturbed, and they need psychological interventions rather than a police intervention or a force. If you have a hammer, everything looks like a nail. If we're given data about threats and given the tools to deal with threats, then we’re going to treat everything like a threat.

**Molly: Thank you! Could you briefly touch on how you got to ethics of care as a framework? What makes an ethics of care framework particularly well-suited to replace models of threat as an approach, particularly in predictive policing?**

Dr. Asaro: First, it's relational, so it's not just a single dimension of metrics. We're trying to improve a particular trajectory, action, outcome, or efficiency on one kind of dimension, but we have to think holistically, and that's challenging. It's very easy to say, “this metric is too low, we’ll need to improve it.” But we really need to think about all these things that we normally don't think, “what are these implications of making a certain transformation in a technology or putting a technology into a different kind of piece of society?”

Aimee van Wynsberghe, who works in robot ethics, has looked at ethics of care with her dissertation in care robots. You can think about the operations of the hospital, and you're trying to maximize delivery of care, improve patient outcomes, and you can measure that in various ways. But when you start trying to maximize efficiency in those parameters, you miss the point that a lot of what happens in care work is care, and that it's not just how much medicine you get and how accurate that is, but that there's bedside manner. There's treating people like human beings and respecting them and their humanity. Nurses and doctors know certain ways that computer programmers don't. We can start thinking about that when we do design from the very beginning, and then wind up with better systems if we do. Especially in policing, or any other kind of care, education, medicine, a duty of care [is] expected. Teachers are expected to take care of students and doctors take care of patients. Lawyers also have duties of care to their clients. It's often very difficult to articulate, but it's an always-present moral duty or obligation.

Some people talk about ethics of care as a virtue ethics. I think that works if you think, “what would the virtuous caretaker do in a certain situation?” There's also community ethics, look at what benefits communities. Most western ethics is very individualistic, so that's problematic because we're building social technologies, not individualistic technologies. The whole history of engineering ethics is about the moral responsibility of an individual engineer: build a reliable system, not approve things that aren't safe. But actually, it's not just them and their moral character that matters. It's the whole society or sub-community that's impacted by a system.

So Western philosophy isn't very good at that at all. Other philosophies [are]. Ubuntu and African philosophy is very powerful. I was just at the Social Studies Science Conference in Hawaii, and they have a very powerful indigenous philosophy, which is one of abundance. I think it comes to this model of threat, which is Western philosophy. You can just read right out of *The Republic*: “all these other city states are trying to invade us and steal our stuff, we need an army, we need a police force, and we need to invade, steal their stuff.” That's the basis of thinking about interrelations, whereas in Hawaiian philosophy it’s that we live on an island of abundance. As long as we take care of the island and we take care of each other, there will always be abundance, instead of thinking about it all as scarcity. Models of threat was about scarcity, and if that's the foundation of your philosophy, then you're always going to [think], “what's mine?” versus “how do I care for others?” If we're all in this moment kind of caring for each other, we're probably better.

**Andrew: In addition to your work on predictive policing, you're also very well known for your work on lethal autonomous weapons. In light of our discussion of models of threat and ethics of care, in what ways are your work on predictive policing and lethal autonomous weapons connected? More broadly, how are the development of both technologies connected in a material and a philosophical sense?**

Dr. Asaro: I think the obvious thing is the connection between the potentials for violence and weapons. I have another set of papers on police robots, and particularly armed police robots, and why those are a terrible idea. All of the justifications that we give to police officers to use violent force or lethal force aren't really acceptable at all for robots, because mostly it's about self-defense, and that robots don't have right to self-defense, because they're not selves. But even in defense of others, it doesn't make sense in most of these situations where you're trying to protect the third party to actually introduce a lethal or armed robot in the situation. A threat [is] the intention to do harm to somebody, as well as the capacity to do harm to somebody. So if I have a weapon, and I'm using it in a threatening enough manner, that creates a threat. But a robot would have to both understand enough physics—not just recognize an image of a gun on camera, because guns could just be lying on a table—and then also understand social psychology and our actions enough to understand this person is threatening this other person.

So they would need a very robust understanding of the physical world, and the social world, and if they have that, then they also should have all sorts of other ways of intervening on that. And this also goes back to the models of threat because we justify or permit police to use lethal force with a gun in all kinds of situations where there's probably other options to de-escalate. But because their lives are at risk, we've written laws that say it's okay for them to use lethal force with very low standards of “they just have to feel threatened.” That means they go straight to this tool or option that is lethal in its first instance of use, like a gun. Whereas, if you understand as a robot or as a person the complex social interaction, the complex physical interaction, you could intervene, you could de-escalate socially, talk somebody down, convince them not to use force. Or you’re a robot: they have a gun, but you could put yourself in front of the gun. You can interact with the physical world or the psychology and social interaction to remove the threat. You should try all of those things before you try lethal force, but there's an expedience to the gun, so that winds up being socially permissible. But we shouldn't transfer all those morals for humans onto machines because they're not humans.

And similarly, with military things, it's an extreme case because there's a lot of violence that's permissible, including civilian casualties as long as they're not intentional, which is also rather broadly construed and difficult to enforce. But ultimately from an ethical perspective, and even in just war theory, the justification of killing is highly exceptional. You can only kill enemy combatants who pose a threat and are still fighting, and if they surrender you can no longer kill them. If they're injured and can no longer fight, it's illegal to go and execute them.

So it's not just *carte blanche*, and you can't just kill your fellow soldiers, that's still murder. So there's actually a lot of rational justification that needs to be in place before killing is permissible. Robots, automated systems, just don't have any access to that. They're not moral agents, they're not legal agents. They cannot justify in making a choice to kill, and it's impermissible morally for us to delegate that kind of decision to them, because we're abdicating our moral responsibility by allowing the machine to make those decisions.

So that's part of that connection, and it's related now to my work here at the Center for an Informed Public. I'm looking not just at violence and threats of violence, but deception and coercion in AI systems and chatbots, and how manipulating you through selected information also undermines your autonomy as a human being. We shouldn't allow machines to do that. How do we actually regulate, I think, is a much harder problem. I thought it would be really easy just to ban killer robots, because it's kind of obvious, but it's been thirteen years and we're getting pretty close now. But hopefully, we can catch up with these chatbots.

**Andrew: You suggested in your predictive policing paper that we should be shifting to an ethics of care approach in the design and the implementation of these technologies. But a particular focus of care ethics literature, as you mentioned, has been an emphasis on human relationality and caring for and about others. What are the limits of an ethics of care and applying that to these technologies? Is it even possible to encode or apply an ethics of care to lethal autonomous weapons? Or is there a deeper moral distancing that's inherent in their design?**

Dr. Asaro: Short answer, no. I don't see killer robots enacting an ethics of care, and I think that's why they should be banned. I don't think there's any ethical standard that we could program into them under any kind of ethical system that will make it acceptable. And so we should just prohibit [them]. I think that the bigger question is, how do you implement this in or with systems. And I think that's a more challenging and important question, because it's more about humans deciding systems: their moral approach to that, as well as the evaluation of that system, the ability to revise and change systems when they're shown to cause harm. All those mechanisms need to be in place, but also from the initial design, start thinking about those things.

And it's not that we're trying to create an autonomous agent that's going to care. I think that's maybe feasible when we have superintelligence: they become moral agents, deserving of respect and part of our society. But right now they're tools and we treat them as tools, but they're tools through which we interact with each other. I'm a media professor. Everything is media. We're mediating our experiences and our relations with each other through all kinds of different technologies. And this is a new, very powerful technology that uses data in complex ways. But fundamentally, it's a tool. And what we need to be thinking about is how we're caring for others through the design of it, through the use of it, through controlling the kinds of data that are in it, as well as setting up this kind of regulatory procedures and mechanisms, laws to ensure that they're doing what they're supposed to do, and are making society better hopefully for everybody, and not just a select few.

So all of these questions about participatory design, I think, are highly relevant. But they're also a little bit misleading in the sense that I also don't think engineers and designers alone have all the responsibility for what these systems do. They're incredibly complex. Even a very simple technology, putting it out into the world, you don't know what people are going to do with it. You try to make it safe, and you try to show them how to use it to benefit society. Ultimately you have to rely on society, rely on users to do good things with it, but you also still have some degree of responsibility.

**Andrew: You seem to be suggesting that the similarity between predictive policing and lethal autonomous weapons, and their tension with care ethics, is a result of their nature as weaponized tools. To what extent is this tension with care ethics specific to weaponized applications of AI like autonomous weapons, as opposed to paradigms of categorization more broadly?**

Dr. Asaro: I think it is very general, hopefully. That was trying to plant a seed for other people to think about how to apply it to other domains. But I think there's obvious connections and ways to do that within weapons. And in general, the way that we think about international security or national security, or even policing. We tend to fixate on threats and not necessarily the underlying problems. I hear this a lot about killer robots: “Wouldn't it be better if robots fought the wars because they wouldn't make any mistakes?” No, not really. They're going to be a lot more efficient at doing things, but that's also going to make it much more likely to go to war because leaders are going to think they're really reliable. You've told them they have this really great targeting AI system that's not going to kill any civilians, which isn't possible. And then you say, we were totally responsible because we put out these things that were designed not to kill civilians, but they killed a bunch of civilians, so that's not our fault. It provides rationale and justification for it, and it's also a way of avoiding dealing with underlying problems: the political issues of the war but also the responsibility to train soldiers or police officers for de-escalation. We get fixated on this one form of lethal force or law enforcement, when a lot of what soldiers do is community relations: digging wells, helping reconstruction. Police do a lot of community engagement, making people feel safe in their community, if they're doing their jobs well. If you fixate on threats, this whole system of militarization is bled into policing.

**Andrew: Your point about lethal autonomous weapons licensing violence relates to another paper of yours where you talk about the relationship between lethal autonomous weapons and totalitarianism. Does the mere acquisition or development of lethal autonomous weapons imply a shift to this more totalitarian form of power? In other words, is it a contradiction to speak of a democratic government that has taken up the use of lethal autonomous weapons?**

Dr. Asaro: When we think about technology in general, and AI and automation technologies in particular, a lot of what they're doing—you're increasing automation, you're increasing efficiency, but you're also redistributing power, and a lot of that has to do with labor. We've had authoritarian and totalitarian regimes for centuries, millennia, but they've always required people. A leader on their own, if nobody follows them, is actually not very powerful. Where power comes from is in training all of these people to do what you say and believe that your authority is real. Hannah Arendt has written about this in *On Violence*, thinking about totalitarianism and particularly police violence, mostly reacting to the student demonstrations during the Vietnam war. If authoritarian rulers had killer robots, they would have this tremendous new political power, because right now, as we understand authoritarian regimes, you need secret police, thugs, informants, and a surveillance system in order to be an authoritarian ruler which means you can ignore public interest or public opinion.

But if you can automate that, you can reduce the class of the police or the number of elites that you have to have around you in order to maintain power. I think what we're seeing now with these mega-billionaires wielding enormous amounts of economic power and media power and political power [is] that this automation is also going to enable ever greater distances of inequality, but also concentrated power in a smaller and smaller number of hands, which is, I think, fundamentally anti-democratic. Not to say that the secret police of a traditional authoritarian regime are super democratic, but even Stalin had to appease a certain number of elites to stay in power. Now, I think these technologies will reduce the number of people in those circles.

**Molly: I'd really love it if you could dive in a little bit deeper to your current research and what you're doing with the Center for an Informed Public, with AI and chatbots increasingly affecting our visual information ecosystems and social media. You've explored ethical concerns regarding these algorithms that strategically manipulate users through targeted content. What ethical issues do you foresee? What impact on our autonomy do you foresee with new technology, like eye tracking technology or generative AI, and how those technologies are designed to improve the persuasiveness of targeted content?**

Dr. Asaro: Sure, lots to cover under all of that, but let me just give the highlights. Part of it relates to something that's in that predictive policing paper, because I talked a little bit about the concept of pre-crime. If we think about targeted marketing and advertising, and how it functions, it's population statistics, essentially. We gain certain pieces of information about you as an individual, we map you into this demographic population model, knowing a few features that define you, and we can predict lots of other features of you, or things that we might be able to sell you. For political interests, knowing who your friends are and how you feel about the certain set of issues to project you, they know how to target you with different types of political messages. This is very powerful compared to traditional modes of advertising persuasion where you're really trying to come up with much more general kinds of messaging for mass communication, because you're sending out a single message to everybody, or maybe you would have a certain subsets where you knew you could get a certain kind of target demographic. But now, you can address an audience of one and that can be very powerful and needs to be regulated. A lot of that's going to depend on privacy regulation and making sure that companies either don't acquire the kind of data that's needed to do that, or if they have it, that they're not permitted to use it in certain kinds of ways to manipulate people.

But I'm also worried about what's next, because historically, the reason mass communication worked the way it did is because it didn't have access to all that data. You have really pathetic psychological models, and it makes lots of wrong assumptions about you. They're actually really bad at it, and they don't really try to build an individualized model of your psychology and what you desire and hope for. They're just still fitting you into a one-size-fits-all population statistic. But now they're collecting so much data they could build models about you and figure out what you care about and who you care about and who you listen to, and they could fake messages from those people or convince you those people believe these things to get you to believe something, or threaten those people and try to coerce you into all kinds of things, and really start manipulating your understanding of the world in a highly customized way. The potential is there for it, because now we have data and the computational power. They just don't have models, but they could start building them and improve them over time. I think that’s incredibly dangerous to think about all the different range of applications that might apply.

Particularly within democracy, it's challenging because we actually value persuasion and public discourse. And that’s a lot of what's happening in the debates on social media. Freedom of speech is a good thing. But it's also pretty obvious that, and it has been for a long time, that speech is not equal, and certain people on those platforms have incredible power over others. You can look at the number of followers as roughly equal to the amount of power. Power announces itself in that way, but it's also real that they can attract all these people, and they can get their followers to exercise very complex types of social threats. It’s kind of new, also not that new. Go back to *The Republic* and Plato [says] the problem with the public square is rhetoric. People should use logic, not rhetoric. Rhetoric was just persuading people using made up arguments. That's not right, we need truth. So we've lost truth in a lot of ways, an epistemic grounding and reliability of our communication systems. I don't know how we restore that, but I think that's going to be crucial. But a lot of that just depends on the public, and if we're constantly just manipulating everybody going forward, how do you get to some system where you can trust it? That’s complicated.

**Molly: In the context of democracy, we spend so much time on systems that are ruled by algorithms and then take that into our worldview, our identities, and our belief systems. Looking back to previous election years and the immense swaths of data that were collected by Cambridge Analytica, how do you see things like generative AI impacting the future of our democracy and the direction that it might take?**

Dr. Asaro: There's a sense in which these generative AIs create things that are plausibly human. They’re useful for that, because you don't need a real writer, you can just give simple instructions. But I think a lot of political persuasion at this point, and even manipulation, really depends on having some kind of insights about society and politics, and at least insofar as you give the system a prompt. Maybe you're generating better messages that way, at least initially. Current generations of these chatbots, I don't think, are going to be super useful any more than sock puppets have been for getting your social media to trend and get things in front of people. But you could just make up a few messages and just replicate them everywhere and get them trending, then that's the key to reaching people.

I think it's these more sophisticated models that are going to be much scarier. And again, Cambridge Analytica—for all of the pomp and circumstance that they claimed this powerful mode—they gave people these really basic personality questionnaires, and they mostly sucked up all of their data from Facebook. What they were doing, as far as I can tell, was identifying persuadable voters in swing states, and that was their value add. It wasn't that they really knew how to convince those people of anything, but they said, “these are the 50,000 people in Michigan that you need to send targeted ads to, because everybody else in Michigan's already made up their mind.” So it has that power, and it can persuade enough people to vote in a close place that matters, but in a broad sense they didn't convince the whole country of anything, and they probably didn't even identify what psychological factors would be influential on those people, merely that they're the most likely undecided voters that could be persuaded in some way. But that could get a lot better in the future because they had really crap psychological models, and they didn't really know how to do any of that, but they didn't need to.

A lot of what AI gets applied to is fast, cheap solutions. Not necessarily cheap—you have to do a lot of computation—but it's fast, and it's cheap in terms of labor. All these automation systems do is reduce the cost to be able to do something. So it takes a long time to learn to paint, but now you can just tell a computer program “paint me a picture that looks like this or that” and it spits something out. The other thing from a purely information theoretic perspective: the level of information in a message is inversely proportional to the likelihood of receiving the message. So a message that you're expecting carries very little information. It's a message that you're not expecting that actually has a lot of information. But what these systems are literally designed to do is generate the next most likely token or word, so it's actually providing the least informative thing, mathematically speaking, that it can at every instance—they're generic generators. So it can be creative, it can be unexpected because you don't know how it works, but it's just generating the very most likely thing that it can, which in that sense, it's not going to be creative. It's just going to find latent connections between data at best, which can be really useful, because there's a lot of data. But I don't think it's going to be super brilliant anytime soon.

**Andrew: That's really interesting because like you said, on one hand these large language models are producing very expected results, but their emergence has also been unexpected in many ways and has caused people to raise questions about the nature of consciousness. What do you think that these technologies reveal about the nature of human relationality, to bring it back to our discussion of care ethics? And do you think that they necessitate that we rethink any assumptions about human nature or about care?**

Dr. Asaro: No? Well, yes, of course. I think if we go back to this idea I introduced earlier about sociotechnical systems, what was really innovative about ChatGPT is not some massive technical innovation, it's that they put a really nice graphical user interface around it. And the thing about Altman is not that he has some brilliant insight into language or AI, he's a really good marketer. He's the Steve Jobs of AI. And Steve Jobs didn't really have any great ideas. He went to Xerox PARC, and they had great ideas. So then he figured out a product that everybody could interface with really effectively and used things that were already out. But that's important, because really, technologies are sociotechnical systems. So you need these marketing people to promote the social side and understand how the technology can effectively integrate into society, and how to convince businesses that they need it and sell it and make lots of money. The actual technology hasn't really changed much in these large language models. They do really brilliant things, and we're all very impressed by them, because now we can actually interface with them in certain ways. How long that enchantment lasts remains to be seen. We already know they hallucinate. They're terrible at rules: they can't do basic arithmetic, but also rules that we might care about like logic, they can't do causal reasoning, they're not going to learn ethical rules.

Now, there's a degree to which they're modeling all these statistical patterns within written language that has been scraped and fed into the machine. What this really is is a giant compact statistical model of all the stuff that gets put into it—that's all a neural network is—but it allows you to access it really, really, quickly to answer queries or to generate text. This idea of predicting the next token as a for generative AI is really fascinating, and it tells you all kinds of weird and interesting things about the texts that have been put into it. I don't know what it tells us about us other than we're the people collectively who generated at some point all those texts.

I think when we start talking about consciousness, it kind of irks me because it's nothing like consciousness. It’s not even trying to be. Some people argue if it were just embodied and engaging in the physical world, then it would just learn all of that really fast, and then it would be conscious. But the first thing is you can't do that with robots. Robots and AI are very different: robots are much harder to program because there’s so many more bugs—they’re a nightmare, don’t go into robotics unless you really love robotics. AI is way easier, because everything is just data, and it's so much faster to fix bugs, to do iterations. AlphaGo, DeepMind’s Go playing computer, is playing trillions of games of go not only against every known recorded game of Go, but also against all of itself as an adversarial network trillions of times to develop the kind of skill it needs to beat the human. You can run those simulations trillions of times. Run a robot around this room a trillion times. How long is that going to take? The sun is going to explode before you finish that. And that's just this room, much less a robot that could deal with the world outside. I don't see that happening. I think embodiment is crucial to consciousness.

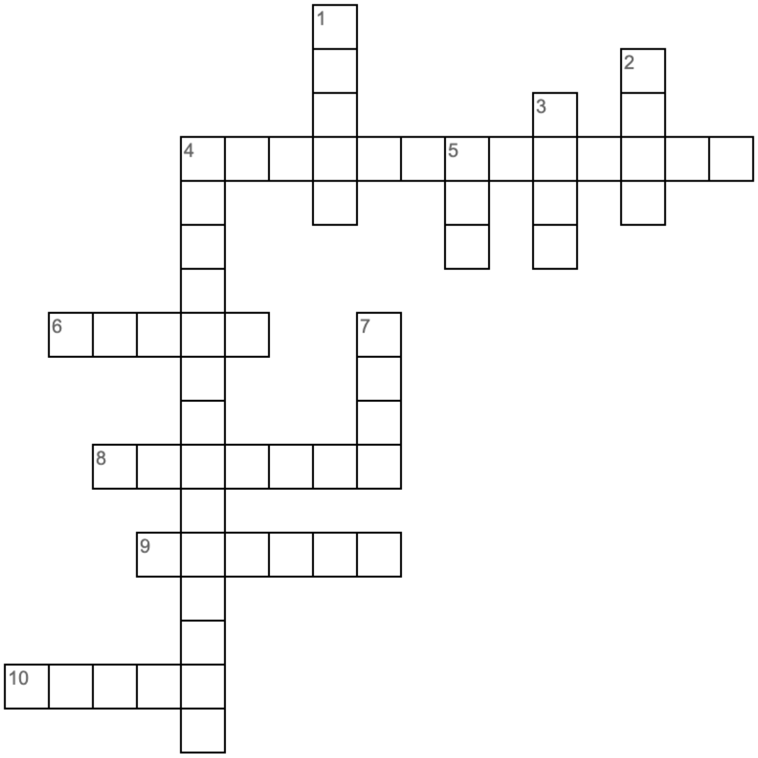
And actually, consciousness is cheaper in a lot of ways, or easier than intelligence, as we understand it in our language. Goldfish are conscious, right? They interact with their world in a conscious way. They're not going to write literature anytime soon. They don't need to. They just need to swim around, find food, reproduce, because they're fish. So that sounds like consciousness is more about a relationality of being able to understand your environment and relate to it. And we do have systems that are getting something like that, with SLAM (Simultaneous Localization And Mapping) in robots and drones and self-driving cars, that are starting to look like they can perceive a three-dimensional world, interact with it, and understand their relation to it. But it's still very limited and very brittle, and they're nowhere near as conscious in that sense as a goldfish at this point. And even if they achieve goldfish consciousness, they’re not going to take over the world. We're not worried about goldfish taking over, right? Being able to interact socially or politically is so far off.

What we're worried about in morality or ethics is, “should these things have rights?” I think that comes to questions about the conditions of having rights and participating in society. That's about both having responsibility and the moral and legal responsibilities of being a member of society that you then incur respect from other members of society as equals, in some sense. They would have to do a lot more than have consciousness than even superintelligence to have what's required for that. They would have to be members of the society in the right way. And maybe if some alien superintelligence and for outer space, instead of a computer, we wouldn't automatically think it's a part of society. We might fear and respect it because it's an alien intelligence.

But I think we also anthropomorphize all of this stuff way too much, and thinking that it's thinking, that it feels anything, that it's emoting, that it's anthropomorphic in some sense, or other fears around superintelligence, that it's going to take over the world and enslave us, we're projecting like how we behave towards other people. We're afraid of those things, and so we think the system is going to be like us and do this to us. But again, if an actual alien comes here, they're going to be so different from us, we probably won't be able to fathom that. Movies make them always humanoid—although *Contact* was pretty good, because it's just totally different.

# A PHILOSOPHY PUZZLE

FOR YOUR ENJOYMENT



**Across**

4 Some difference in X is necessary for a difference in Y

6 Ethics of care in predictive policing say what?

8 Chewing on universal grammar sounds like nom nom nom

9 existence precedes essence

10 Ethics bowl host, and a student of Aristotle

**Down**

1 wrote an essay while inhaling nitrous oxide

2 Categorical imperative

3 Best departmental advisor award goes to...

4 beyond the call of duty

5 Neglected

7 This neuroscientist knows every fact about color perception but not what the color red looks like

ANSWER KEY: