Rationally Speaking #220: Peter Eckersley on “Tough choices on privacy and artificial intelligence”

Julia: Welcome to Rationally Speaking, the podcast where we explore the borderlands between reason and nonsense. I'm your host, Julia Galef, and it is my pleasure to introduce you to today's guest, Peter Eckersley. Peter was until recently the chief computer scientist for the Electronic Frontier Foundation, which is a non-profit focused on promoting privacy and free speech and autonomy on the internet.

Now he is the director of research at the Partnership on AI, which is an organization that includes a lot of the major tech companies, focused on developing best practices around artificial intelligence as it develops. And his PHD is in computer science and law at the University of Melbourne.

Peter, welcome to Rationally Speaking.

Peter: Thank you Julia.

Julia: So, you focus on a number of topics including privacy, and artificial intelligence - and safety and regulations around artificial intelligence. So we're going to cover a lot of that, but I thought we'd start with privacy.

There has been an increasing amount of attention, public scrutiny of ways that tech companies have been failing to protect our privacy, especially since the Cambridge Analytica scandal. I was wondering what you think about how well the public's attention on this issue is allocated? Do you think that we are basically most concerned about the most important privacy problems — or are we overreacting to some things and under reacting to others? What's your take?

Peter: I think it's hard enough to answer the question theoretically for ourselves as experts about how important privacy is. It comes ... The way we think about privacy comes from a very animal place. I think of it as being: You're out on the savannah around a campfire and then when you see eyes in the darkness watching you, that feels really dangerous and really bad. And I think that type of psychology is the mechanism that's at work amongst the people who care a lot about privacy.

They therefore are very cautious about what they share with whom and then try to get control of the way that technology starts to collect data about them. But of course, that technology is so complicated that it's really hard for most humans to have any notion as they using an app or a phone or a website, what the real implications of data that they're sharing or it's being revealed about them invisibly by the devices might be.

I think we struggle psychologically with our animal reaction and this very complicated technological world. And then you add an extra layer over the top, which is one of the societal consequences of failures to protect privacy in various ways. And those seem to be really very diverse. You can argue about how important they are, but they include things like:

In the United States, where the criminal justice system is massively overreaching and incarcerates on the order of millions of people who shouldn't probably be incarcerated. This is a consequence to privacy violation, which is that it leads to more arrests and more imprisonment. And if you do some math on that, it looks very serious.

There's a different consequence in people's personal lives, when their partners or their families learn things about them, and then have problematic power relationships with those people. Maybe you don't want your conservative family to know that you're trans or gay. Maybe you don't want your partner who's domineering to know about all of your social life, et cetera. And those are very high stakes problems that certain people face. They're very different from the criminal justice case.

And then I think in your comment about Cambridge Analytica you were getting at the third and perhaps craziest example of this. Which is that we built the web the first time around to reveal almost everything that people were reading and thinking all of the time to websites and third party tracking companies. And there was this theoretical concern about this, but we finally saw it come home to roost.

I have a story to share about this. Maybe 10 years ago I worked at EFF, and I had this amazing colleague, Cory Doctorow, who people may have heard of, he's both an activist and a science fiction writer. And he came to me and one of my colleagues at EFF and technologists and said, "I want to write a story about how Google turns evil." This is like 2007, maybe 2008. What would Google do if it were evil?

Julia: That is a great generator for a story — like, that general kind of question, what happens if this thing turns evil? What would that look like?

Peter: That's right. In fact, everyone you should all just pause your podcast right now and go in and think about your own answer to this question.

Julia: But come back to the podcast.

Peter: Yes, if you don't come back, you won't get to hear the one that I gave Cory, which he turned into a short story actually. So the answer I came up with was, "Oh, Google would mess with politics." It would figure out how to swing elections and totally gain control of the political system.

Julia: Peter, how sure are you that your idea wasn't the inspiration for Google turning evil? It seems like an information hazard, potentially.

Peter: I don't think you should just assume that you caused really large effects in the world by answering a question from Cory Doctorow — though Todd's ruled that out.

Julia: I was just last night looking through lists of inventions that were inspired by science fiction so I'm primed to suspect you in this case. But anyway, go on. So Cory wrote the story...

Peter: In fact, he worked the story a little ... My suggestion was, well, Google could read all the candidates' email and know how to understand the motives of all the humans, and kind of mess with their campaigns will help them.

But Cory's version was, "Oh, it's the way that the candidates are perceived by the public that's totally shaped by such results." And so, if they want to help you, they'll show you all these great search results about you. And if they don't, they'll surface sordid things, perhaps even fictitious, sordid things about you.

Now it turned out it wasn't Google, that completely led the charge on this, it was Facebook, and it wasn't deliberate. It perhaps happened by accident to a large degree. So there are maybe certain people at Facebook who had some notion of what was happening, but we wound up in that world. And I have an apology, which is after that conversation I didn't do anything to stop it.

Julia: Great. So you're outlining various categories of consequences of what happens if privacy isn't ... Is the example of Google, hypothetically, or Facebook actually swaying the results of elections, or swaying public opinion on various topics — Is that about privacy? Or... I mean, it seems like without collecting a bunch of private data they could still decide to curate your news feed to prioritize things that are favorable to one party versus another. How does that relate to privacy?

Peter: Well, I think the dimension that turned out to be privacy oriented here is ... what would be driven by a lack of privacy is that the platform sees what you're interested in and engaged with. And he can A/B test the user interface, and then do machine learning on each human specifically, to figure out, well, what type of stories is Julia interested in in particular? And it then turns out to be way more effective to radicalize people around propaganda if you can customize that messaging to each individual person.

Julia: Right, yeah, that makes sense.

Peter: Interestingly, also, I think Google has been accused of having done this in certain ways as well.

Especially on YouTube, where the reports are that the related videos or suggested videos tab over on the right hand side would start... If you started with a reasonable educational video, it was really easy to wind up clicking through a series of videos that went through conspiracy theories, all the way to radical Islam or the world is flat. And these things are just optimized for, "Well, it seems like someone who watched this thing might watch this next thing and be engaged by it."

And it turns out that documentary content making arguments for those things can be really compelling. And so you just showing people a series of documentaries that tells them that they live in a world that's flat.

Julia: Interesting. So, I've been trying to think about the forces working against privacy — which in the case of the government might just be a, increasing power, b, wanting to increase security. Like, the desire to catch criminals or terrorists is in conflict with the privacy of citizens in a lot of ways.

And then for companies it seems like the forces working against privacy are the desire to optimize ad revenue, and then also the desire to just make their services work better for people. In order to give you an experience that you enjoy and will keep coming back to. And I guess we could certainly argue about whether apps that people keep coming back to are in fact apps that are good for them… But still, apps that people want to use.

In order to make that happen that requires, or it's helpful to have, a lot of personal data about people and their usage habits.

Do those seem like the tensions to you, the forces working against privacy? Did I miss anything?

Peter: Those are the big ones. And there are some ways in which we could imagine addressing some of these and others that seem more profoundly [intractable]. We're just being confronted with the fact that it's really convenient to give up privacy to giant technological platforms, and then really hard to verify that they're actually working in their interests.

Julia: So, which of those anti-privacy forces feel most tractable to you?

Peter: Conceivably … I don't want to say super tractable, but conceivably the advertising revenue incentive could be addressed.

Julia: Like by subscriptions?

Peter: For instance. We could … more more people could choose to buy products from companies like Apple. And part of me is very uncomfortable saying this for other reasons, because they're a very closed proprietary technology company. But they do a slightly better job of representing the interests of their users, because their users are paying them a lot of money for those services, whereas other platforms will tend to be free.

But it's not just subscriptions and paid models that we could look at. Maybe the United States will never do this, but it's well within European, Canadian, Australian tradition to have public funding producing amounts of important media content and supporting authorship. Every time you borrow a book from a public library in most countries in Western Europe and Canada or in Australia, the government will pay some amount of money, maybe a dollar to the author, sometimes the publisher of the book.

We could look at models like that to decrease the necessity of relying on advertising, and therefore very intrusive advertising, for technology makers and authors and content producers, new sites. I think it's a little bit of an unfashionable idea in the era of neoliberal economics, but we should seriously consider those options, at least encouraged our European friends to seriously consider them.

So you can imagine a way past advertising incentives to spy on everyone — but the piece where it's just more convenient to us, that's hard to fix.

Julia: Yeah. So do you see part of your mission — I mean you're no longer at EFF, but your mission is just, like, generally someone working for a better internet. Do you see a part of that mission as just convincing people that privacy is more important than they think it is? And that it's not worth making the trade off of more convenience for less privacy?

Peter: I don't think convincing people is necessarily the way I thought of that. More just making sure that they've heard the argument. Here's what ... Here's the choice you're making if you choose to not run a tracking blocker or an ad blocker on your browser. Here's the choice you're making if you share information on Facebook.

You might still make those choices, but make them informedly.

Julia: Yeah. I tend to think a lot about ways that our reasoning and decision making faculties are not as perfectly adapted for the modern context, or our goals as individuals. As opposed to our genes' goals, or our ancestors goals.

And it seems like one big category is little decisions, each one of which is not that consequential, where the costs are not that noticeable or consequential — but over time they add up, and probabilistically make us a lot worse off in the long run. But that's not very salient or, or visible to us each time.

So this is seems like why we have trouble saving money for the future, or dieting, or whatever — because on the margin, each choice... The benefits of eating that piece of cake are very salient and immediate, and the costs are probabilistic and indirect, and in the future.

And so giving up pieces of privacy feels very much like it falls into this difficult-to-reason-about category.

Peter: Absolutely. One of the things I would often find myself saying from my position at EFF was: There would be much more of a market for privacy if you could buy it in retrospect.

Julia: That's a really good point.

Peter: If you realize, "Oh, I just lost my job," or "I just was subject to this form of social ostracism" - or identity theft, in the extreme case — and I can time travel back now and take privacy decisions differently... I think people would take more of them.

Julia: Going back for a moment to that distinction that I made between the incentives of governments to try to push the boundaries of privacy, or violate privacy, on the one hand, and the incentives of companies... Are you more concerned about one versus the other? Do you think the consequences of companies versus governments being able to violate people's privacy are greater?

Peter: We obviously have historical models of situations where governmental violations of privacy were so severe that they probably sit at the top of the list. If you lived in Eastern Germany or the Soviet Union or these other totalitarian states... In those situations your entire life and liberty were at stake all the time as a result of

privacy violations. And people in those countries had to think about sharing their political views even in private with their partners, for instance, because their partners might be informants.

And so in a setting like that, the technologically assisted version of that… I think one way of thinking about it is it couldn't be that much more intrusive than Eastern Germany, but it could be much cheaper for nation states in the future to be that intrusive. East Germany was effectively spending a huge fraction of its GDP on monitoring its people. And I think we'll see other states do that much more cheaply with video camera surveillance and phone monitoring and email monitoring, et cetera. China is probably the laboratory of far you can go with that, but maybe other states as well.

And so I think that there's a really very high variance, if you ask how important are these effects in the United States. They look pretty large because of the criminal justice problems. And then there's another term in the equation for, what is the risk that we wind up in a totalitarian version of the United States. That's extremely bad.

Maybe there are enough institutional protections that we're going to hold off trends in that direction when they happen. I think it's a year when people probably wouldn't gamble super aggressively on that claim. Those failure modes look really bad.

If we look at the corporate failure modes, corporate, private privacy failure modes, for the most part, they don't grow as extremely out in the bad direction. They are not as catastrophic — except for this swinging elections thing, which then turns around and feeds into governmental failure modes.

Julia: Yeah. I want to ask you to do another comparison now — between, not industry and government, but instead between different tech companies. So Facebook, Amazon, Google, maybe Twitter, WhatsApp… which of these companies do you feel are doing a pretty good job of protecting their users' privacy, and which not so much?

Peter: I get cautious about trying to make categorical comparisons between these companies for a few reasons. They often make different things and then if you look inside that product lines, even when you can find comparable products, you'll see places where each of them is doing the best job in a specific way and then doing a poor job on other fronts. I think of them almost more as being like countries that have their own problems and governance structures and they get certain policy areas right and certain poly policy areas wrong.

I can give random examples, like — having praised Apple before, I'd say one thing that Apple really doesn't do right is iCloud backups, which contain almost everything on your IOS devices. Those are totally readable by Apple and disclosed to government in response to law enforcement or surveillance requests.

So all of the FBI drama about getting into a San Bernardino iPhone was kind of theater because the FBI had full access to the very recent backups of everything on that phone.

Julia: Well wait, why did they bother with the theater then?

Peter: Because they wanted a legal precedent that said they could compel Apple in the future.

Julia: Oh, interesting. And... did they choose that case because it was an especially bad dude, and so they had more potential to...

Peter: They'd been a dozen that they've chosen not to use before the one that they chose.

Julia: Quite interesting.

Peter: And so in that specific product direction ... Android recently launched an encrypted backup feature that's much stronger than iCloud backup. So for now on that specific access, Google is doing a better job, but we can name probably half a dozen other places where that's not true.

Julia: I ... this could just be my imagination, but I feel like some of these companies are just like sketchier or less trustworthy than other companies. Like I get a bad vibe from Facebook and I don't get as bad vibe from Google, for example. And that could totally change. But in terms of how hard I think they're actually working, behind the scenes, to do what they say they would do — that that seems like it's a company-level variable, that differs between companies.

Peter: I think that's especially true in these companies that are still significantly run by their founders, and where the personalities and inclinations of those humans have shaped the way the company operates. Like in the case of Facebook, Mark Zuckerberg has a history of having done things himself that are like not super encouraging on privacy, or a “responsible exercise of power that his platform gives him” front.

Julia: On the other hand-

Peter: Now the question is, how much has he listened to all the interventions that have been targeted at him?

Julia: Right. I was going to say, I saw all these photos of him feeding cows in Oklahoma and putting his hand on his heart in Baptist churches across the country. So surely he is a changed man.

Peter: I'm persuaded.

Julia: Switching tacks a little bit, I assume it must be the case that our laws in the U.S. on privacy are kind of woefully outdated, just in the sense that they were mostly written before the era of big data and the internet. There have got to be a lot of ways in which they're not appropriate for the current landscape of privacy risks.

Would you agree, and if so, what do you think are some of the main mismatches where a law is just not appropriate for the current world?

Peter: Just jokingly, I mean — I think maybe I would disagree with the premise that there *are* privacy laws in the United States.

Julia: That's way worse than I thought. Wait, why do I think there are privacy laws?

Peter: There are some, if you break it down by government versus commercial privacy. In the commercial space, there's very little. There's enough law to say that if a company writes a privacy policy and then violates it, companies are required as a result of Californian law to write one. But it can be a vague motherhood statement -

Julia: Did you say “motherhood statement”? Is that jargon?

Peter: Yeah. It might be an Australian English term. Like you know, a general reassuring declaration of something.

Julia: Nice, like, as if from a mother.

Peter: "We care deeply about your privacy." Facebook cares deeply about your privacy. Twitter cares deeply about your privacy.

Julia: Yeah, I think it's statements like that that actually make me distrust companies more. Every time Facebook even gives me notifications that are like, "Julia, we care about you and your friendship," that kind of thing... Every time I see that, it just downgrades my trust in the company a little bit.

Peter: And then they'll say things like, "We may collect data, including the following."

Julia: Yeah. Okay, yeah. So companies are required to follow what they say they'll do, but they have the freedom to word what they say they'll do in a very vague and non-binding way.

Peter: And then if they violate those policies, which is already — there's no reason for them to have done that, they could have written permission for themselves to do the thing they wanted to do — you can't sue them. You can go to the Federal Trade Commission, which is a body in DC with limited resources, that will triage and investigate a small number of these instances, and then potentially in some cases fine a company. But those fines typically are very small compared to the amount of revenue that's at stake for the companies.

Julia: So their only incentive really is just PR backlash? Is that it?

Peter: Yeah.

Julia: Wow.

Peter: And essentially the reason the FTC seems significant is largely because of the PR consequences of being investigated and fined by the FTC.

Julia: Would you write new privacy laws? I could be misreading you, but you seem like generally wary of heavy-handed regulation as a rule. But in this case it seems like it might be necessary to preserve people's freedom. How would you balance that?

Peter: I probably would have written a rule that says people need to have a way to really opt out.

Julia: What might that look like?

Peter: We spent a lot of time trying to make this happen. It could have looked like a setting in your browser or in your phone that says, "When you turn this on, people you have no relationship with can't just give themselves permission to track and surveil you. You have to specifically agree." Basically, something like the regime that GDPR has now created.

Julia: GDPR?

Peter: Yeah, in Europe. Would have applied if you chose to opt out. Instead, it's this weird situation ... We're probably not going to have enough time to dive into all of this. It's this weird situation where Europe — that was significantly more willing to impose serious penalties, 4% of the company's revenue, for failure to comply with its privacy rules — that will get to write the privacy rules for the world because of that legislative willingness.

In the United States, there were — especially in the Obama years — a whole bunch of bills in Congress that were trying to create some basic rules of the road. But they were never going to get 60 votes in the Senate to pass.

Julia: Why do you say that Europe will get to write the world's privacy laws?

Peter: Well, if you're a company and you want to do business on the internet, it's a lot easier for you if you can reach all of those European uses.

Julia: Is this a good thing, as a rule? It seems like maybe in this case it's good, but the generalization of this is that whatever country wants to be most restrictive, as long as they're a moderately big country, that sort of forces the rest of the world to follow those same standards.

Peter: It's weird game dynamic. I don't know whether it's good or bad. I think it can be played out in constructive or less constructive ways in different spaces. As a person working on policy or doing activism, you always want to just look at the game board like that and say, "Oh, if I get California to do something, I can shape U.S. policy," or, "If I can get Europe to do something, I can shape world policy."

Julia: I thought I read something actually on the EFF blog recently about Article 13 in the ... I forget what the overall piece of legislation it was in the EU that would require large platforms to have a database of copyrighted images, and censor things on their platforms that violated —

Peter: This is terrible sloppy thinking on behalf of Europe. I mean, a lot of this thinking came out of the impact of the internet on news organizations, traditional media organizations. And the internet has really massively shrunk the amount of revenue that's available to those companies for somewhat subtle reasons, right? There's just a lot more places to put an ad, and Google controls many of those places. So the ads in your newspaper are no longer able to command the same premium that they used to.

Then the classifieds in your newspaper all disappeared and went to Craigslist or a local competitor — but in response to that, the newspapers, especially in Spain

said, "We want to be able to control what Google shows in Google News." Which is, if you look at Google's products, that's a tiny little piece of the stuff that Google does, and it's probably ... we don't know, but it's really quite small, revenue wise. So Europe focused all this regulatory energy on, "How can we control what Google News displays from a Spanish newspaper?" They managed to get Google News shut down in Spain by trying to extract revenue there.

Article 13 I think is trying to generalize this terrible theory of how European media outlets can claw back some revenue from Google. I think there are a lot of other ways of approaching that question that could be more constructive. Maybe we do need to turn to the tech companies and say, "How are you going to fund a healthy media landscape?" It should be not tied to specific absurd copyright claims.

Julia: Is this the kind of thing that would force changes across the rest of the internet, because tech companies have to conform to the law in the EU, and they can't selectively conform?

Peter: It depends. There's always a cost associated with building two sets of products and having them work differently in different places. Sometimes the companies look at the cost of complying with a regulatory regime and say, "That's so annoying to us that we're either going to pull out of that country, or we're going to go to the trouble of building a different product there and not having it be the same thing that we do everywhere."

Facebook did that with the GDPR. They just said, "Okay, there's a version of Facebook that's GDPR compliant, and there's a version that's not, and everyone else gets the non-compliant version." A lot of other companies will just make everything GDPR compliant.

We might see with Australia and the UK seriously considering mandating back doors in encryption, companies having to choose: do they put a back door in for those governments, or do they just stop offering their more secure messaging products and all the associated features in those countries?

Julia: Zooming out a little bit now —` how have your views on privacy and the landscape of trade-offs and strategies evolved in the years that you've been working on it?

Peter: Actually, there's one thing I wanted to add just before I answer that, which is all of the things I said about privacy law are about commercial privacy law. There's a separate area of government privacy law where in the United States, we do have the Fourth Amendment, which in theory, secured people in their homes against search and seizure by the government. Those protections have been massively eroded by the war on drugs and the war on terror, and so we still do nominally have a Fourth Amendment, it's just far flimsier and less useful as a protection against various forms of, whether it's law enforcement surveillance, NSA surveillance, et cetera.

Julia: But actually — to jump in before you answer my own question — I was thinking when I asked the question about, "Are our privacy laws mismatched with the current landscape," I was thinking of things like email. Which people now use as

they used to use letters, and there were laws that you can't open someone else's mail — but your employer can read your email. It seemed to me they weren't subject to the same kinds of strict privacy protections that letters were, even though emails are now the modern letters.

Are there things like that that you're concerned about, or is that just a trivial example?

Peter: Workplace surveillance is an entirely third category almost from the two that I was talking about. Because it's almost more like a family relationship, where you have this very close relationship between one human and a surrounding group of humans, and the surrounding group of humans gets to create a lot of norms and conventions. And in most places the law will back employers up in setting policies that require surveillance of an employee's communications.

Whether that's actually good... I think a lot of us feel it probably isn't, a lot of the time. It does depend on the nature of the work and the nature of the roles that people have. I feel we haven't gotten the right outcome in a lot of settings, and a lot of people have workplaces that are pretty psychologically intrusive as a result. Some of that is about the law, and some of it's about bargaining power. It's also an area where I'm honestly less of an expert.

Julia: Okay, so going back to the question of, have your views on privacy and laws thereof changed in the years that you've been working on it?

Peter: I think, weirdly, I found myself at first working on privacy because I'm one of those people for whom privacy is an emotionally salient topic.

Julia: Eyes in the night.

Peter: That's right. I get really bothered by the idea that people can see things on my computer, but I didn't necessarily think that the topic was as important as my emotional reaction to it. I thought that other areas of internet policy, like copyright law and access to knowledge, were probably much more important.

Until one day I did a back of the envelope calculation on this question and realized, "Oh, there are some places where privacy potentially in certain societies just becomes way more important."

They're all a little bit indirect. It's never that privacy itself is the most important thing, it's just that sometimes it's a safeguard against something else terrible happening. That then raises this weird question, which is how strong can that safeguard be, especially in an era where technology's made privacy invasion so easy?

Julia: Are there any significant disagreements or debates within the community of people — EFF and its sympathizers? Are there any disagreements within that community about privacy?

Peter: I think that a huge one that people have been grappling with in the last few years is the relationship between privacy, freedom of speech, and the political character

of online debate, and the consequences of platform design. There are a lot of places in there where you see a trade-off between privacy and say competition.

Julia: Competition - how so?

Peter: If you wanted to encourage the creation of alternatives to Facebook, the first thing you would do is require Facebook to make a lot of APIs that let people get their data out, let competitors to Facebook user's data out so that you can build something else. Those APIs are exactly the same sorts of APIs that Cambridge Analytica was able to exploit to [harvest] vast amounts of user data.

Julia: Wait, why is it a threat to my privacy to give me the right to extract my own data?

Peter: Unfortunately with social media networks, or social media [platforms], it's not enough to get your own personal data. Much of the data that's there is things that relate to interactions with other people.

Julia: Oh, I see. Right, so I can't just extract my half of a conversation that I had with someone, because at the very least, you know that I have the conversation with that person, even if you block their comments, and... yeah.

Peter: Exactly. One of the big pieces there is an address book. Can you get an address book out?

Julia: Oh yeah. Right, because then it shows that they know me, not that just I know them, so a commutative relationship. Right. That's tricky. I mean, do you have a position on that issue? Or is your position, in a nutshell, that it's complicated?

Peter: This is a personal one, it certainly in no way represents the view of any organization. Actually, that's probably true of everything I've said today.

Julia: Okay.

Peter: For me, I think that the political risks of having massive privacy invasion tied to machine learning algorithms for politics are really high, and so I might err on the side of trying to get the platforms to protect privacy more strongly over competition. But I'm not super optimistic that we can win that either. I might be being irrational in thinking that this is the thing we should do even though we can't accomplish it.

It also contains this question of, okay, so I suppose you get Facebook to get better at protecting privacy, which means keeping all the data in a box itself, and only using it itself. How do we get Facebook to use that data in a way that's democratically constructive? That's a super hard governance question.

Julia: Yeah. All right, well maybe that's a good point at which to shift over to discussing AI, which you've been working on ... well, for a while, but especially now that you've moved to the Partnership on AI.

I was just looking at an impressively detailed comprehensive analysis that you did recently, of progress in AI in different domains of AI, from image recognition, to speech, to transfer, and so on. Can you talk a little bit about: A, does that project

relate to your concerns about privacy, et cetera, or is it a totally separate thing? And B, how do you go about measuring progress?

Peter: On the relationship, it definitely has informed some of my concerns and helped me to understand where we're at on political implications of AI. The project that we did, we wanted to ... this is an EFF project, you can find it at eff.org/ai/metrics, the AI Progress Measurement Project. The idea was to just get some high level sense of: When we start as a policy organization working on AI, should we focus on these short-term issues that are obviously relevant, like the use of machine learning prediction systems for sentencing and pretrial detention in the U.S. courts, which are responsible for ... the machine learning algorithms are responsible for probably 500,000 people being incarcerated.

Should we focus on that? Or should we be thinking about the way that artificial general intelligence at some point might totally transform the planet? What evidence is there or isn't there, that that type of scenario is imminent?

The way we did this, we just picked a methodology and tried it out. We said, "Let's make a list of problems that we know humans are able to learn to solve, and then sub-problems under some of those problems. Then for each problem, do we know any metrics or ways of measuring where the machine learning systems are yet able to do this thing, learn to do this thing?"

We just made a list, and we got order of couple of hundred of different sets of metrics for order of 100 problems. I should have those numbers in front of me, but it's roughly 100. Then we went out and looked at the literature and said, "Okay, so how well are the current neural network architectures solving this reading comprehension problem, or this game playing problem, or this vision task?"

What we found was progress across a lot of fronts has been very fast, especially in the last five years. There are almost three buckets of problems. There are the ones where progress has been really rapid, and humans have already been surpassed in basic vision tasks, and Atari game playing. In a handful of very basic reading comprehension problems, we're starting to see human level performance, but those are second grade reading comprehension tests.

There's a second category of tasks where the progress is happening, but it hasn't reached close to human level yet. Things like answering questions about what's going on in a photograph.

Julia: And that doesn't count as image recognition?

Peter: It's in a fairly advanced image recognition problem. So the basic task of, "Can you tell me, is there a microphone in this photograph?" Or, "Is there traffic in this photograph?"

Julia: Ones that we keep helping the algorithms solve every time we answer captcha questions.

Peter: Exactly. Those look solved. The problem of, show a picture to a neural network and get it to either write us a correct long form sentence description of what's

going on there, called captioning, or the problem of be given a picture and a question, and answer the question about that picture — So, this is not just a vision task, but a hybrid vision and language task. Those problems ... that one's called visual question answering. Those we see rapid progress, but you wouldn't really say they're solved.

Then there's a third category of things that are just beyond the current state-of-the-art. We don't yet have anything that looks like a real neural network writing computer programs of a general sort, or answering questions about the behavior of computer programs. There are certain advanced kinds of language tasks, where you're really wanting to not just answer simple questions, but kind of get the deeper narrative out of a story.

For instance, some of those tasks look too hard for the current state-of-the-art ... Sorry, reading scientific research papers would be another thing where you're like, "Yeah, maybe in specific domains you can do some feature extraction, but being able to critically read a complex text is way beyond the state-of-the-art."

Julia: I noticed this kind of disconnect when I talk with people about AI progress — between the people who are quite bullish, just enthusiastic about the amount of progress that's been made so far, versus the people who are kind of unimpressed.

I guess there's two sources of the disconnect, and I'll say both and you can just respond to both of them. One objection from the, let's call them the bears, is they'll point at specific examples of an AI giving a really dumb answer to what should be a really easy question. Like showing it a photo of some sheep grazing on a hillside, and the AI's like, "That's a lady in a hat." People post stuff like this on Twitter, and they're like, "Yeah, I'm really not worried about a robot uprising." So that's one type of objection.

Then the other is a more general, "Yes, we may have made a lot of progress on these domains that we can measure, but these domains seem like very narrowly defined and pretty small pieces of the totality of what would constitute human intelligence." So like, I don't know, being able to perform really well on this specific video game, or being able to accurately recognize images of things that you were trained to recognize with a certain set of images, like of dogs, or sheep, or whatever.

It often seems like the debate is ... the bulls will say, "Look, we're like 10 times better than we were two years ago," or something, and then the bears are like, "Yeah, but we went from like half a percent of the way to AGI to like 2% of the way to AGI. Yes, that's 10 times better, but we're still pretty far."

I've been trying to resolve this debate. And it's tough, because there's no easy objective measure of what percent of the total way to general intelligence does this type of image recognition constitute?

Do you have any thoughts on ... if you're bullish about progress, then can you address why these specific narrow tasks seem like an important cause for optimism?

Peter: One thing I will say is that if you go to that webpage, eff.org/ai/metrics, you'll find a really deliberately hard to engage with, incomprehensible document. It's vast and hard to get a big picture from it. A thing we thought about, and might still do, but we thought about the idea of making something much more simple and digestible. The cartoon version of this could be a progress bar. You know, one of those Windows or whatever style things, where you can see we're 70% of the way there-

Julia: Yeah, and it could go backwards as we learn more about how hard the problem is.

Peter: That's right, it would be exactly like one of those Windows progress bars where it gets to 99% and then just sits there and spins for 15 years. So for various reasons, I think people are cautious about the possible over-interpretation of that, and then a little bit also cautious about information hazards around making it super clear to everyone where we're at.

The intention of the project was to have the source materials for that view. And so, having worked on it closely, I kind of have an answer — which is I'm moderately bullish, and doing that project made me more so.

The reason for that is, to address those arguments from the bears... These tasks are specific, but the interesting thing is that it was not that you learned to play a ... it was not that you built a system that could play a specific Atari game, or specifically play chess, or specifically play Dota2. It's that you were able to teach a general purpose system, like a reinforcement learning agent, or some supervised neural network architecture, to do this thing.

There is some genericism in the progress. And you see papers that pop up that make significant progress in wildly different sub-fields of machine learning, simultaneously sometimes, which is very interesting as well. It looks as though there's a generality to the types of neural networks that we have.

That may not be sufficient to get us all the way to AGI, but we can kind of ... we're starting to get some confidence that intelligence can be made from things like these deep neural networks.

I think that's one reason for bullishness, and the other one is just that the progress that's happening is on so many fronts, and it's continuing. It's super dangerous to try and extrapolate any kind of line out from that — but if you do, and then you wave your hands a lot, you get something like 15 ... you could convince yourself that 15 years of this trend would close down the current list of problems that humans can learn to solve and machines can't right now.

Now of course, you then should update to expect there to be new ones and hard things you hadn't anticipated, so maybe you wave your hands and then you go from 15 to 25 years.

Julia: Waving one's hands is a underappreciated forecasting technique.

So — you worked in computer science and law. I know very little about the laws around ... like IP laws around machine learning algorithms. Have people been

patenting these algorithms? And what role does patent law play in encouraging or slowing down progress?

Peter: Well, the precedent we have is from the last 20 years of the software industry where we saw massive patenting of all sorts of ideas and simple algorithms, complicated algorithms in computer science. The total consequences of that giant wave of patenting seemed to have been very negative for the software industry. I don't think there's total consensus, but there's sort of near consensus that these patents get stockpiled both by the companies that actually make products, and by these other types of entities, non-practicing entities — or patent trolls, as they're more accurately termed.

What those companies do is they just extract money from people who are trying to do useful things, either small companies or large ones, and then run off with it. So they're a tax on the software industry. There's no evidence that the patenting process actually causes invention. In software it seems that people invent things in order to accomplish goals, almost always. It's very rare to have the kind of, "We're going to pull lots of money into this abstract R&D thing to make an algorithmic step forward."

Probably the closest you could get to that would be video file formats, and other things where there's just a vast amount of work that goes into making an MPEG file what it is, but those are so exceptional. Almost all of the work that programmers and computer scientists do is towards a specific thing they're trying to build right now, and they solve the problems that are in their way.

As a result of that, and probably a lot of other dynamics I can talk about that make software a little bit different from other fields of invention, the patents have been a massive problem for that industry. It's struggled and tried to get significant patent reform, to shield itself from some of these problems — but for various reasons, both political and dynamics within companies, that didn't succeed. It hasn't succeeded yet.

So the prior we kind of would have from the software industry is that machine learning would be awash with patents in the same way, and it would wind up with the same problems. We've seen it awash in the patents, and I have not heard as many reports of trolling and the huge problems we saw with the software industry and machine learning yet.

Julia: That's interesting, why?

Peter: I only have wild speculations about that question.

Julia: Sure, that's what we're all about here, is wild speculation.

Peter: One wild speculation is that it might not be as easy to tell when you have a good target for a machine learning patent shakedown. I mean, I haven't gamed that out. The other dimension might be that there aren't as many products that are obviously machine learning products. The machine learning in tech products is often quite subtle — though that's not a very good explanation, because though ordinary people using them might not realize when machine learning is happening, I think experts probably do.

Julia: Yeah. I mean, can companies be forced to share the algorithms that they used in a particular product?

Peter: During a lawsuit, potentially.

Julia: Yeah, yeah. So you could look at some other company's product and be like, "I bet they're using the algorithm that we developed," and then you could sue them.

Peter: Actually, the fact that there's guesswork in there ... Well, it depends, right? So in some cases when the model, the neural network model ... Is everyone familiar with that term? I'm just trying to listen to the audience. They all know that a model is like this term for a trained neural network that does a specific thing.

Julia: Uh-huh.

Peter: When the model for, say, recognizing your friends in your photos — when that thing lives in the cloud, like on Google's computers, or Facebook's computers, it can literally be quite hard to get a copy of it. Potential patent trolls might really have trouble knowing how to sue those companies. Whereas if the functionality lives on your device and works offline, so if your phone can recognize your friends in a photo when you're not connected to the internet, then in that case, potentially if someone roots the phone or jail breaks the phone, they could extract the model and figure out, "Oh, it works this way."

Julia: Cool. I guess last question about AI progress before we wrap up: have you been thinking at the Partnership on AI about what kinds of regulations might make sense? Either official laws to regulate potential downsides of AI, or unofficial agreements between tech companies about how they're going to ... like what kind of safeguards they're going to put in place, or ethical practices to follow?

When I talk to people about work on AI safety and related topics, a big source of skepticism is like, "Yeah, in theory it would be good to have safer AI instead of less safe AI. I agree there could be some risks, but it just seems so futile to try to think about regulating a technology that doesn't exist yet." I mean, if we're talking about general intelligence, not just image recognition. And there aren't historical precedents for this, et cetera. How do you feel about that?

Peter: I think it's too soon to regulate general intelligence, absolutely. But in specific domains of machine learning and AI, there are absolutely open, high stakes, urgent regulatory questions.

Julia: Like what?

Peter: I can give you four examples.

Julia: Sure.

Peter: California at the end of August passed a bill called SB 10, which started out as being probably a very constructive reform to the criminal justice system. It abolished the money bail system that is responsible for large amounts of incarceration in California. At the last minute, that legislation was amended to mandate that every county in California should purchase criminal justice risk

assessment tools, which are essentially machine learning tools that predict whether someone's low, medium, or high risk, and then — not in a completely deterministic way, but — use those tools to make decisions about whether people are incarcerated prior to trial or conviction.

We know from excellent research from ProPublica, and following ProPublica from various academic groups, that those existing tools have massive bias problems. They're hugely disparate in their labeling of especially African Americans and other minorities as high risk. That's a giant problem, and the question is, will this bill cause the propagation of these biased tools in California?

I think we're thinking there's a hard question about what to tell to the judicial council. It has this review body under this legislation about this new regime that's being set up. Is there a way to get other standards that can be applied to tools to ensure that you're not using ... you're not purchasing tools that are massively biased in every county in California, and then deciding 60,000 people's fate with them?

That's one example. Another example, which we looked at at EFF, were proposals to mandate labeling of bots on the internet. Those proposals have a lot of failure modes to them. They can accidentally wind up forcing platforms to label a lot of humans as bots, because they have to make a decision in a hurry, and the standards are super vague, and there are risks and very high incentives to not fail to label a bot as a bot.

The legislative drafts also didn't distinguish between bots and what I might call a cyborg, or a hybrid between a human and a bot. Of course, many systems are not clearly either human or machine learning, or chatbot.

Julia: They're not clearly, in the sense that we can't tell? Or that they literally are a combination of a human and a bot?

Peter: Well, if you use the Gmail app on your phone, you'll notice that it starts to have suggested replies. If you just click one of those, was that your voice or Gmail's voice?

Julia: Right, but presumably if Russia or some company is going to be using bots on Twitter, say, the whole point is to save time and do it at scale, so you're not going to have humans doing this stuff.

Peter: Well, on the contrary, what you want is for the bot to be really effective, and so you build a giant lookup table of conversational paths that you've seen before, and that you know what to say next for, and then whenever you get a question, or a comment, or get to a situation that you haven't seen before, rather than making up an answer that is probably going to be terrible, and will reveal you to be a strange, inhuman robot, you're headed to a human and say, "What should I say next?"

Julia: Well, since we're almost out of time, why don't we leave it at two examples, because I'm itching to ask you my classic end-of-the-episode, Rationally Speaking question, which is: Can you nominate a resource, like a book, or blog, or even a

person — an author, thinker, whatever — that you have substantial disagreements with, but nevertheless, think is valuable to read or engage with?

Peter: I was thinking about this question. There are many different directions to try and answer it in, but the one that I think I'm going to share is this book by David Graeber called Debt: The First 5,000 Years. It's an absurdly ambitious, certainly overly ambitious kind of grand narrative of history, and of the way that markets displaced the preceding cultural human economies that existed in tribal societies, and the inherent violence in that transition process and its implications.

This book tells so many great yarns about so many dimensions of life, and it appears to back them up with a lot of strong citations, but I can't believe that it's all true. In fact, I can't believe-

Julia: Have you done any spot checks, epistemic spot checks?

Peter: Maybe one or two, and they've been mixed. So what I almost really want is to have a Wiki-fied version of this book, where we go back and say, "Okay, of all these beautiful claims, how many of them really check out, and how many of them turn out to be something else altogether?"

Julia: Oh, nice. I do have a friend, I'll link to her blog, she does epistemic spot checks. It's just her personal blog, it's not a paid thing, but — she'll take a book that makes a bunch of claims, and just pick randomly 10 of them to check or something, and use that as a barometer for how trustworthy the book is. Which is a thing I wish was more widely done.

Peter: That's an excellent idea. I'm sure a lot of your audience has read Steven Pinker's Better Angels of Our Nature.

Julia: Mm-hmm.

Peter: I read those two books together, and I found the experience of reading them together to be really striking.

Julia: How so?

Peter: Well, it may be my politics, but Pinker really annoyed me.

Julia: How?

Peter: He just seems so wildly overly optimistic, and seems to just really ... and overconfident is maybe the biggest thing. He's like just, "Here's how it is. I'm here to tell you that violence is not a problem anymore." There are little subtle weaknesses in his argument that make me very skeptical of it.

Julia: What's an example?

Peter: So his numbers on warfare are all about battle deaths, and he just excludes deaths from war outside of battle, and says, "Well, they must be declining as well in proportion, but it's hard to get data about them." It just seems like oh, that's obviously not necessarily true. It could well be the case that modern warfare

causes many more proportionally speaking non-battle deaths because it ranges much more widely because of mechanization, for instance. Or maybe that's not true — but it's the kind of doubt that should suffuse the book, and it doesn't.

Then the thing that's striking about both of these books is they address a lot of topics that seem to be off topic, you know? They'll verge into the topic of honor, or witchcraft, or-

Julia: Because it's like thematically related, or aesthetically related, or ...

Peter: Both authors seem to have claims about how these other concepts fit into the main subject that they're arguing about. The thing that really struck me reading them together was that Graeber and ... Pinker is making this claim that violence is declining, and in particular, tribal societies had huge problem with violence. Then what Graeber does is goes through and tells specific contrasting anthropological stories about different tribal cultures, and how they created norms that caused or mitigated violence.

You get this picture, "Oh wow, this was really complicated." There clearly was a huge problem with incentives to violence in hunter-gatherer and early agricultural societies, but culture really responded to that incentives in some really creative ways in some places, and not everywhere.

Julia: Does that contradict Pinker's thesis though? It sounds like it could totally be the case that culture adapted to respond to the problem of violence, but not nearly enough to mess up the trend that Pinker's pointing out.

Peter: I don't think it necessarily contradicts it, I just think it's grounds for a lot more caution than Pinker engages in, like a lot more of a complex narrative than Pinker seems willing to tell.

Julia: Well, this is fun. It's making me think that I should add or substitute the Rationally Speaking question to be, "Please criticize a work that we suspect many of our listeners will be a fan of," because that seems like a good practice.

Peter: Or mix it up each time.

Julia: Or mix it up, yeah. Shot and chaser. Cool, well, we'll link to both of those books, as well as to blog posts and other articles that we've discussed during the episode, and to EFF and the Partnership on AI. Peter, thank you so much for joining us on Rationally Speaking. It's been great having you.

Peter: Thank you, Julia. That was a lot of fun.

Julia: This concludes another episode of Rationally Speaking. Join us next time for more explorations on the borderlands between reason and nonsense.