Tech Summit on Artificial Intelligence: A Quote Book

Data and Models Edition

April 2024
US Federal Trade Commission
Office of Technology

Table of Contents

Themes

A Positive Vision for AI-enabled technologies

Governance grounded in human needs

"It would mean having an information governance regime that thinks about information beyond being property. It thinks about information, and the harms that arise due to privacy invasions, to labor violations, to consumer rights violations." - Cory Doctorow

"This current AI race is based on certain assumptions about both scale and speed as a proxy for progress. And it's a view that's based on narrow benchmarks, it's one that never really properly contends with the longer term environmental, or labor impacts, or the impacts on our information environment." - Amba Kak

"[W]ho gets to decide and shape what counts as innovation, and what counts as innovation for the public good? I think that one way forward is to really go back to the drawing board, or the table, which is currently populated with VCs, big tech firms, and companies that they invest in, and really have a much more broad ranging conversation that is dominated by public, rather than very narrow private interests about what counts as innovation and what is innovation in br4on anda(I)-3.2 (d)-10.2 (i)2.3 (a-5.8 (b

Potential consumer protection issues related to AI model development

Direct consumer harms

"Although we described some very valuable things by calling them property, that the most valuable things in the world we describe with non-property language, and that's people. Harming someone is not theft of their integrity. Killing someone is not theft." - Cory Doctorow

"People really do feel a lot of pressure to race to get things out there." – Jonathan Frankle

"If it's possible to pay for the data, and still enact the same harms, still displace creative workers with the work that they've done for you, still possible to produce grotesque privacy invasions in the form of non-consensual pornography, still possible to harm people by mining their data to make inferences about them that are adverse to their interests, then we have managed to fail to solve the

(n)-5.6 (.1 (3.2 (c)-h-5.1 7(ti)et-2.3 (m)-1.p)-2719 T-1.9 (r)y. Td( )TjEMC /LB/TT.8 (eo)-5.42.51 Tw 09 Tw 28.18 0 Td[

which a small number of firms have enormous amount of bargaining power over the creative workers who generate the value for them." - Cory Doctorow

"So I think a question for the very near horizon is to try to wa

tech giant goes up and tries that as well, that their investment just won't be able to survive." - Stephanie Palazzolo

Misaligned incentives

"Having the money to pay for data licenses is not correlated strongly with being someone who will not harm the public, that there are lots of incumbents, with lots of money, who've got a strong track record for being the last people we want to lead us into the future." - Cory Doctorow

"I think the pitch to investors is that we are going to tell hospitals you can fire half your radiologists and double their output, and that is not the AI productivity and benefit world that we want. That is the alignment problem that we're worried about." - Cory Doctorow

"...there is nothing that is inevitable about the current trajectory of AI. That is really important to keep remembering and reminding everybody. Because I think this current AI race is based on certain assumptions about both scale and speed as a proxy for progress. And it's a view that's based on narrow benchmarks, it's one that never really properly contends with the longer term environmental, or labor impacts, or the impacts on our information environment." - Amba Kak

"We are nowhere near a place where a bot is going to steal your job, but we are well beyond the point where your boss can be suckered into firing you and replacing you with a bot that fails at doing your job. And I think that's the real AI alignment problem we should be thinking about." - Cory Doctorow

Access to talent and capital

"I've noticed a lot of investors going after companies that are founded by ex-OpenAI researchers or maybe scientists that were at Google or from very... some of the top colleges in the U.S. And that kind of makes it harder for founders that maybe come from other types of backgrounds to get funding and to get capital from these investors." - Stephanie Palazzolo

Data advantages

"Big tech firms have a very clear advantage here from the last decade of commercial surveillance." – Amba Kak

"As a related point, we also don't know if and to what extent these data advantages will port to the so-called AI startups that they are strategically investing in, potentially creating new forms of dependency and power asymmetries outside of those that already exist via the compute and cloud arrangements." - Amba Kak

"I think one of the biggest misconceptions I typically come across in our field, and especially when I chat with folks in policy is everyone just assumes that because they know of OpenAI and ChatGPT, that's the only business model and that's the only way of operating." - Jonathan Frankle

"Even getting the data to do this is incredibly expensive. One of the most important expensive inputs to building these models isn't compute. It's the data hand labeled by humans to make the model good at specific things. And for any of the fledgling startups that are out there, that three to six months, and those millions of dollars probably have to be weighed against racing to market, and making a name for yourself in a really competitive environment. And that incentivizes risk-taking. So that's certainly one of the consequences of competition that I imagine is on the minds of a lot of people I know at small startups." - Jonathan Frankle

Inability to sustain a competitive foothold

"'How are these businesses sustainable in the long run?' And the question of business model, and how startups that are making open source models make money is very important if you're one of their customers that's depending on them to build a product on top of their open source model. And

Brightline rules

"Data minimization isn't new at the highest level. It's been around for more than a decade globally in various forms, in various laws including the GDPR. So I guess, 'Why hasn't it worked,' I think. Or, 'Why hasn't it prevented some of these, the worst privacy invasive practices?' I think there, the lesson, if anything, is one on not allowing too much room for interpretation. Because I think where maybe the first decade of data minimization came to a head was on the question of, 'Is behavioral advertising a legitimate business purpose? And if it is, does that mean we can just maximize, collect as much data, and keep it forever?' And I think as we look forward, acknowledging those administrability challenges, acknowledging that an interpretive wiggle room will be abused, to really focus on bright-line rules that don't allow that, that make it very explicit that AI training is not a free card to break down all your data silos, to violate purpose limitation, that we want to draw bright-lines around restricting particularly the use of the most sensitive data like biometrics or related sensitive attributes." - Amba Kak

Integrate competition and consumer protection

"[P]ushing for more integrated regulatory approaches that don't silo out the consumer protection side of things, and the competition side of things. Because we have, again, seen how some of the largest firms really took advantage of that over the last decade to amass the information asymmetries that they have, and further concentrate their power." - Amba Kak

Strengthen laws

"To address these harms, we have to reach to things like labor and privacy law, not copyright law. It is not enough to merely have the right to feel affronted by conduct of firms. We should have the right to do something about it." - Cory Doctorow

"I think Americans often underestimate just how primitive the state of American privacy law is. The last time we got a really big muscular improvement to our national federal privacy regime was in 1988, when Congress got worried about video store clerks leaking their video store rental history, and passed a law prohibiting that activity. The Internet's come a long way since then." - Cory Doctorow

"If you are worried that TikTok is making millennials quote Osama bin Laden, or if you're worried that Facebook made your Grandpa a conspiratorialist, or that Instagram is making your kid anorexic, or that protesters at Black Lives Matter demonstrations, or the people who attended the January 6th riots are all being identified by Google through reverse warrants, you are someone who cares about privacy, as is anyone who is worried about the privacy implications of AI, whether that is models memorizing, and then regurgitating private information as we've seen, where sensitive information from medical histories, or resumes, or commercial databases of purchase history are sometimes being memorized and coughed up by these models, or whether you're worried about the truly grotesque generative AI problems with image and video generators, things like non-consensual pornography generation, copyright's not a great tool for dealing with this. But privacy law would certainly give you an awful lot of remedies to deal with." -

Integrating disciplines for decision-making

"I think there's just so much room for professors and researchers at these universities to be part of this conversation. And they don't have the same incentives that profit driven companies do. And I think we really need to be encouraging that a lot more."
- Stephanie Palazzolo

"It's not open versus closed source. It's not more laws versus letting people innovate. It's not the academic side versus industry. There's a lot of gray area here, as we all talked about today." - Stephanie Palazzolo

Grounding rights central to human needs

"Merely giving tradable property rights is always going to be inadequate in the same way that we don't solve the problems of a lack of organ donors by creating property rights in kidneys, and then just letting people sell their kidneys. We need rights that deal with our information in a way that is cognizant of, and sufficiently important that we recognize its gravity, and its centrality." - Cory Doctorow

"We need to think about the problems of data beyond a property rights regime, beyond the idea that if you make data, it's your property. And someone else has to pay you, and get your permission before you use it. Because what we want to make sure of is not that everything in the models is paid for, but that the public and other stakeholders aren't harmed." - Cory Doctorow

Improved media framing

"It's the responsibility of, especially the media, and people like me, even though it's much more easier to write stories, and just say, 'Oh. It's X versus Y,' I think it's up to all of us to make sure that we're discussing this, keeping those gray areas and nuance7eSr (thi)-27uharay,- (ay)-5a,e ae-5.8z thi2TJz(l)- (t)-5 70 -17 say,