Stanford eCorner

The End Game of AI

Tristan Harris, *Apture*

05-11-2017

URL: https://ecorner.stanford.edu/?post_type=snippet&p=55063

In the arms race to control our attention, design ethicist Tristan Harris says the latest and perhaps most daunting weapon employed by tech companies is artificial intelligence. The co-director of Time Well Spent discusses how, because of AI's machine-learning capabilities, its algorithms improve with every click, like or other interaction — eventually becoming powerful enough to anticipate, and perhaps alter, our thoughts.

Transcript

- We have this natural situation where everyone's competing for attention, and we are evolving these systems to get better and better at extracting human attention, and we, then it gets really competitive and we have to add something else.. So what do we add? We add AI.. So now instead of just offering a product to you, I actually have to predict with big amounts of data and machine intelligence what's gonna keep you on the screen.. Instead of just offering some stuff you can click on, I'm gonna actually predict from millions of things I could show you if I'm YouTube, what's the video that I can put in front of you? If I'm Tinder, I'm gonna pick from millions of people I could show you, what's the perfect reason to not be with the person that you're with? If I'm Facebook, I could say all the million things I could show you in the news today, what's the perfect thing that's gonna get you to like, click, or share, okay? So you have an AI that's basically been given this goal.. Now, I want you to put this in context.. When we think about AI, we have to remember, you know, when you point the AI system at chess, right, first it kind of wiggles around and it makes some kind of funny looking moves, and then it starts making some smart moves as it gets better, and then it makes some surprisingly smart moves, and then it beats Garry Kasparov, and when it beats Garry Kasparov, it doesn't unbeat Garry Kasparov.. It's now better than all human beings at chess, right? So, you take that same AI and you point it at the Go game and that took 30 years.. It made these funny looking moves and now, with AlphaGo, with Google's AlphaGo, it beat all of Go players, and when it beats all of Go players, it doesn't unbeat all of Go players.. So, we built these AIs, and then we actually invisibly gave it a new target.. We pointed it at this, and we said, whatever gets this human being, play chess against this human being's mind, and play 20 steps ahead of where their mind could even possibly see and show stuff to them that is either that perfect next video on YouTube, the perfect outrageous news story on Facebook, which controls what two billion people will think every day, the perfect reason for you to cheat on your spouse with Tinder, the perfect political message so we can vary 60,000 political messages and we can actually combine word choices and different contortions of politician's faces and colors of buttons to perfectly animate a response from your brain stem, and so we're playing chess against ourselves, and we have every click that we give this system, it gets stronger, right? Every time you click or you share, you're feeding it attention, which feeds it more dollars, which feeds it more resources, which feeds it more computing power, which means it's better at playing chess on your mind..

The reason this should be alarming is that these systems are not neutral.. We have a tendency to say, you know, we've always had computers, smart, I mean, computers, video games, radio, TV.. We always worry about them, so why should we be so concerned this time? And there's a few different reasons that this is so different, and the biggest one is this AI enhancement that I just mentioned.. A couple others are that no other medium could pull on your social psychology, so no other medium could show you an infinite set of reasons of why other people are living better lives than you are, an infinite set of reasons for why you should feel like you're missing out, other people are having fun without you, an infinite set of reasons why you owe people responses.. You didn't open your TV and it said, "You know what, you owe like 100 people responses.. "You better start getting back to them," right? So, we're doing such enormous, I don't wanna say harm or damage, but we're doing, there is so much that is now put on the human evolutionary animal, right, more than we've ever put, and especially when you add in AI.. So now the question becomes, what is the goal of that AI? What does it actually want? So if this is not a neutral product, if this is not just sitting here, but it actually wants something from me, well, if you're Mark Zuckerberg, you think the thing you programmed the AI to do is to make the world more open and connected, which you translate into, "Let's engage people.. "Let's show people whatever engages them," and that's gonna be very persuasive to someone who's living inside of that mind, but we have this problem because the actual thing that this is all based on is attention.. It's the same

contradiction that I felt, and we now have this system that's, you know, we always talk, I don't know how much you talk about it here at Stanford, but there's all of this discussion about runaway AI.. What if in the future we were to build a runaway AI, like a paperclip maximizer, and we give it this goal to make paper clips, and it turns the whole world inside out just to create paper clips, and what if that would happen, and how would we make sure that wouldn't happen? So there's all these people working on AI safety, and the amazing thing is that this basically already happened, because we already built a runaway AI that's steering what two billion people's thoughts are, and we hid it from society by calling it something else..

We played a magic trick on the human mind, 'cause if you call it a news feed, or you call it YouTube recommended videos, or you call it Tinder recommendations, people won't even notice.. That's how easy it is to fool the human mind...