

We all know how The Matrix goes: super intelligent machines trap humans in a virtual reality to use their bodies for energy. What we don't know is where these machines came from. Who designed them? How did it go wrong? I believe that they are just the grand-grand-sons of social media, an evolution of Instagram, Twitter and Facebook.

I first heard this idea in a [fascinating interview](#) with Manuel Cebrian from MIT, I really recommend checking it out. In it, he suggests that these machines could actually be Twitter. See, all of our social media today is driven by algorithms who learn exactly what every user wants, and makes Twitter, Facebook and YouTube a lot of money. The more people use them, the more advertising space they can sell, the more money they get.

So how do these algorithms learn? Turns out that not very different from a dog, in a very Pavlovian way. The algorithms get a reward whenever they show you a post that you engage with, or click on, or look at for a while, that is, if they successfully catch your attention. If instead, you scroll past the article and don't pay attention to it, the algorithm gets a punishment. So effectively, through trial and error, and reward-punishment training, the algorithm learns what you like, in that way, it maximizes the time you spend in each platform. And here comes the ugly part: the different platforms compete against each other for our attention, so all of them are trying to maximize the time humanity spends on their platform by learning exactly what gets users attention. So what would happen if these platforms went beyond deciding what content to show you, but to create content themselves? Think about how many hours of video on stand up comedy YouTube has, for example. It is very foreseeable that an algorithm could be training in the same Pavlovian manner to learn from them, and generate the perfect stand up video for each user. They could even go one step further and not only create videos, but create virtual reality worlds. Since they know you so well, what you like to watch (your Netflix, YouTube history), where do you like to go (your Google Maps history), how the people you like talk (from your WhatsApp conversations), what you like to eat (your Yelp reviews), they could use not-so-distant virtual reality technology to create a hyper-realistic simulation tailored just for you.

You might think, "nah, humans would not want to plug into a virtual reality, they would prefer the real world". But what would people from 500 years ago think of us, staring at a screen all day, effectively disengaging from the real world every 10 minutes? What would they think of the fact that people get home from work and tune out of their surroundings by watching two hours of Netflix? That would be as weird to them, as it would be for us to watch people tune out into a perfect virtual reality, where they have everything they need, want and love.

You might also say, OK, but the machines in The Matrix did that to harvest us for our energy, Twitter does not want to consume energy. Well, actually, the AI algorithms running our social media are incredibly energy-hungry. For example, training one of the latest machine-learning models used around the same amount of energy 60 US households use in a year (Strubell, 2019; US Energy Information Administration, 2017). A self-improving AI, like the ones underlying our feeds, would do anything to fulfill its goal even if that means exhausting all of the

energy in the world. And if that happens, they would figure that they could just use human harvested energy to keep going. After all, why use externally produced energy to feed humans with their attention grabbing virtual worlds, when you can just use the energy they themselves produce and close the loop?

So I would ask: is getting our attention the thing we would be asking our algorithms to optimize for? That is good for Facebook's profit, but is it good for society as a whole? How far could that go? Even if Instagram doesn't turn into a virtual-world-generating-machine that harvests us for our energy, how far should the attention sucking machine go for us to think it's enough? I believe that social media provides value to society, however, the algorithms of privatized social media optimize towards the profit maximization of tech companies and which are in conflict with the interests of society as whole. I believe that social media should be open source, and its code and algorithms developed collectively as a not-for-profit, much like Wikipedia is written or Python is developed. The algorithms in this open-source social media would optimize for other things other than our attention: learning, social interactions in the real world, positive discussions or even happiness of the user which could be assessed through periodic surveys.

Imagine a social media platform that asked you once every couple of weeks how you are feeling. If you select a smiley face, it is a positive reward and keeps showing you similar stuff. If you choose a sad face, it changes its behavior and shows you different stuff until you choose the smiley face more often. There could be multiple objectives, of course so it not only maximizes for individual happiness but other collective values as well. There could also be an option for you to adjust your own objectives, so one week you could ask the platform to optimize towards the amount of time you exercise, therefore, the app could learn to show you content that inspires you, like your friends working out, and articles about the benefits of running for the brain.

Why should the algorithms that control our behavior be in the hands of private companies? Why would the objectives of these algorithms be generating more money for private companies by sucking more and more of our attention? Even though Facebook has pledged to make some changes that optimize towards positive social outcomes, in a profit seeking system, profit always wins. The only way to get socially desirable objectives embedded in our algorithms is to design those algorithms collectively, without a profit in mind.