

# AI and You

Transcript

Special Episode: Disinformation

Episode 69

First Aired: Monday, October 11, 2021

Welcome to episode 69. This is going to be the episode, well the first episode, about disinformation and misinformation. I mentioned that way back in episode one, but I guess that I held off on tackling it until now because it is the hottest of hot-button topics. And I didn't want this show to be defined by that and get cast as the show for either debunking or airing conspiracy theories. I am not above debunking conspiracy theories; we'll do plenty of that. But this show is mainly about understanding artificial intelligence. But since AI has such a role in the spread of disinformation, we can't ignore it on this show, no matter how sensitive an issue it is.

So to start off, right off the bat, so you know where I stand: Vaccines are safe. Bill Gates is not putting 5G chips in them. The moon landings were not fake. 9/11 was not an inside job. John F. Kennedy was killed by Lee Harvey Oswald acting alone. And the Earth is not flat. Plus a whole lot of other things that I really wish there wasn't any controversy about, but you get where I'm going. Now, if my statements on those things caused you any discomfort, well, then I'm afraid the rest of this podcast is not going to make you feel any better. But the spread of ideas like that is what I would characterize as a public health crisis. And proper, thoughtful, transformative, and constructive conversations about what to do about it are long overdue. And that is fortunately now starting to happen, just barely.

A year ago, I was quite down about the direction this was going in, because disinformation is something that has no natural brakes, you know, when something gets so ridiculous, and out there, that people say, "Wait a minute, that's going too far," and they pull it back. Disinformation changes what people think as being too far out. So there's nothing that stops it from going to arbitrary extremes. If you saw some excellent reporting done by the *Wall Street Journal* recently about the TikTok algorithm, you got a visualization of how that works because it showed how the algorithm takes people through watching videos and looks at what they spend the most time on. And they actually, at the *Wall Street Journal*, invented characters, fake users with different personalities and interests, to spend time in front of TikTok, seeing what it showed them. And the more time they spent on certain topics, the more extreme it would take those. And so if someone was naturally depressed, and that was the fake personality that was assigned to this bot, then eventually it would be showing them videos about suicide. And you could see this visualized in a 3D diagram where the people watching would be taken down an arm of this octopus-like shape that showed them getting further and further out to greater extremes of ideology. And that, of course, introduces the AI connection.

It's a similar algorithm on YouTube and Facebook, that they look at what do you spend the most time on? And in YouTube, their goal, the developers, was maximize the time that someone spends on the platform. That's what all of the social media platforms want, right? That you spend as much time as possible on them. So they didn't do a whole lot of psychometrics about their users. They just looked at what do they spend the most time on and used cluster analysis to show them more of the same kind of things. That doesn't mean that there's any kind of conspiracy there with them thinking about taking people in radicalization programs, it just means that the AI has figured out what sort of thing they will spend more time on.

Well, unfortunately, we will spend collectively, individually, more time looking at something not just because it's also cute and makes us feel good, like pictures of kittens, but we will also be engaged with things that make us angry. And so when someone sees their first conspiracy theory, and it's mild enough and plausible enough for them to consider, you know, maybe that's true, then they'll spend more time watching it. And the more they do that, and the angrier they get, the more likely it'll show them something that will do even more of that. And this is what gets them sucked in.

And I'm relentlessly conscious that the more I say this, the more it sounds like a conspiracy theory itself. But I'm not, as I said, imputing any malign desire or goal on the part of the developers; this was a side effect of what they did. But the effect has been documented in numerous ways and we cannot ignore what it's doing to our society. So I'm not going out on a limb there in attributing the algorithm or describing the algorithm the way I did about YouTube. In fact, Sean Parker, the first president of Facebook in a 2017 interview said that his team's thought process when designing Facebook was, "How do we consume as much of your time and conscious attention as possible?" But you can find numerous accounts of people whose relatives and friends have become foreign to them, alien, as a result of this capturing their attention in this process of radicalizing them into conspiracy theory beliefs.

And we need to be careful what ideas we put in our minds. Am I saying that we should avoid ideas? That sounds horrible. But think of it like nutrition for the mind. Your body may be robust enough that you can feed it Twinkies and cigarettes all day, but eventually, that's going to take a toll. And I believe that the same is true of ideas that we put in our brains. So the people who are chanting that vaccines are unnecessary for them because they have "immune systems" are the ones whose mental immune systems failed to protect them against those ideas. We need an FDA for ideas (Food and Drug Administration), someone that will identify the ones that are harmful. I know that this triggers all kinds of reflexes, especially in the United States with the First Amendment, but we cannot ignore this any longer. The First Amendment is not a suicide pact. Neither can we ignore the attractiveness to foreign agencies and subversion groups of being able to weaponize the First Amendment in this way. This, again - it sounds like conspiracy theory. I don't know how to make the words come out of my mouth without that sound, but it is really not controversial that it has been weaponized in this way.

You don't have to take my word for it. Look at Nina Schick's excellent book, *Deepfakes: The Coming Infocalypse*, where she dissects this phenomenon and traces the attribution of where much of the actors are in this. And also, Christopher Wylie's book about what happened with Facebook and the 2016 election, *Mindf\*ck: Cambridge Analytica and the Plot to Break America*. It's not hard to see the motivation that for instance, Russia has, to generate this kind of activity. Now I'm specifically not talking about whether they hacked the voting machines in the election or anything like that; I don't know. And frankly, it's unnecessary because of what they're able to do through social media. And there's plenty of documentation about what they've done in that respect with botnets and people in Russia who are paid to spread disinformation through social media. It's not hard to see the motivation that Russia would have, because destabilization of the United States' influence in the world helps them with expansionist policies.

But let's suppose for a moment, that that wasn't the case, that they didn't do anything. And that when the news came out about Russia having attempted to influence the election through social media, that that was all wrong. What happened? What were the consequences? Were there congressional investigations? Were there sanctions? Did the American people turn against Russia, start burning hammer and sickle flags? None of that. There was no consequence at all. So if they hadn't been doing it, and then they were looking at that reaction, then it would be clear that they had nothing to lose by doing it. If you're Vladimir Putin, and you didn't have a hand in this, then you should buy a lottery ticket, because no one has the right to be that lucky. He's an ex-KGB agent. This kind of psychological warfare is exactly their stock in trade. And how exactly would the United States retaliate? With weapons? With conventional munitions? No, there's a long-standing principle that you retaliate in kind. But Russia is a closed society, information flow is controlled. So the idea that the CIA might be able to launch an equivalent campaign against them and destabilize their government is nowhere near as powerful as what's happening in the US and other Western democracies.

Now, there was in fact, an open US Senate Intelligence Committee hearing about "A Primer in Russian Active Measures and Influence Campaigns" in 2017. And here are some of the comments that were made there by Clinton Watts, a senior fellow at the George Washington University Center for Cyber and Homeland Security:

Russian active measures hope to topple democracies through the pursuit of five complementary objectives:

1. Undermine citizen confidence in democratic governance.
2. Create divisive political fissures.
3. Erode trust between citizens and elected officials and their institutions.
4. Popularize Russian policy agendas within foreign populations.
5. Create general distrust or confusion over information sources by blurring the lines between fact and fiction.

Could you have a more accurate description of what has bedeviled us in the West for predominantly the last five years, but going back further than that? Why? Because Watts said that the Kremlin's goals are "the dissolution of the European Union and the breakup of NATO." And obviously, the 2016 presidential election in the US was only one of the potential ways of influencing that. There's Brexit and then there's what has happened with COVID vaccinations in the West.

And so at a time when we should be coming together to support our authorities and institutions that we allow to govern us, and experts, such as doctors, to treat us and make decisions about how best to protect us, instead, a significant fraction of the population is reviling them, making their job as difficult as possible and causing many of them to quit. So I hope you will forgive me some poetic license when I describe this as World War III, and that we are losing. Why? It is a war, it involves the whole world, and it poses an existential threat. And not just because this has led to tens of thousands, if not hundreds of thousands of excess deaths from COVID in the United States alone, but because of the potential for destabilizing the world order, and the peace that has reigned, relatively speaking, in Europe and the West for 75 years.

And as a reflection of that, on January 23<sup>rd</sup>, 2020, the Doomsday Clock, which has been maintained by the Bulletin of Atomic Scientists for decades now, and represents how close we are to potential annihilation, was moved from two minutes to midnight to 100 seconds to midnight. They now decide to measure it in seconds, so that's 20 seconds closer. January 23<sup>rd</sup>, 2020. Talking about the "two simultaneous existential dangers of nuclear war and climate change that are compounded by a threat multiplier: cyber-enabled information warfare that undercuts society's ability to respond." They went on to say, "The international security situation is dire, not just because these threats exist, but because world leaders have allowed the international political infrastructure for managing them to erode." And clearly, this disinformation has accelerated and caused a great deal of that erosion.

So to make a connection here with AI in case it hasn't been clear all along: why is this such a big issue now? Psychological warfare and propaganda have been tools in war for millennia. What is it that's made them particularly acute now? Think about the state of propaganda and counter-propaganda in World War II. You had people exhorted through posters to watch who they were talking to. "Loose Lips Sink Ships." And one of the other posters from that era, "Free Speech Doesn't Mean Careless Talk." But what were the enemy's vectors for spreading propaganda? The allies are not going to be reading German newspapers. Some of them might be listening to Lord Haw-Haw on the radio but otherwise, it's relatively difficult for that information to cross national boundaries during wartime. Hence the importance of the *agent provocateur*, the enemy sympathizer and possibly paid agent who doesn't take up arms but resides in the allied country to spread ideas that serve the enemy. So they might strike up conversation in a pub, and ask questions about what Churchill is doing and the cost to allied lives - innocent and military. But they can only change hearts and minds one at a time through conversation, even

though it's the most effective way of doing that. But what if you could do that at scale? What if you had one agent provocateur for every member of the enemy population? Well, that's what artificial intelligence offers the possibility of. That you can have a conversational bot that can engage people in that kind of dialogue and plant those ideas.

We're in a transitional era right now, as far as the capabilities of chatbots go. In limited domains, they can hold their own, but once it starts getting more general, they're quickly visible. But we know that the more a chatbot says what the person listening wants to hear, the easier time they'll have of it. Take, for example, the site Ashley Madison, which a few years ago, was basically in business to help people cheat on their spouses. When they went bust, it was discovered that most of the so-called people on the site who were chatting under female identities were actually chatbots. But they didn't have a very high bar to clear. I mean, they just have to start a conversation with saying, "Love your picture. Looks like you work out a lot. Have you always worn your hair in that dashing way?" How critical do you think the men are going to be after that, especially if the chatbot says that it's a woman from another country, and "Sorry, my English not so good"?

Now, certainly the standard of grammar and attention span on Twitter is not something that would tax most modern chatbots. So AI is plainly capable of being the voice whispering in the ear of everyone in a target population. Whether it's been used in that way at that scale yet or not, I don't know. But certainly, much of the activity that has taken place would be in the way of proving out or testing that capability, that potential. And that's all very theoretical.

But as a practical, emotional impact, this has split up friends, and families, and communities. Everyone I know has lost someone they care about to YouTube and other conspiracy theories. So far, trying to do something about this looks like trying to plug the leak in the Titanic with your thumb.

In 2019, Facebook removed what it called a global network of more than 900 accounts, pages, and groups that *WIRED* reports, were allegedly using deceptive practices to push pro-Trump narratives to about 55 million users. Leaving aside the fact that this was the Trump side, for now, the important factor was that these were fake accounts and using fake faces generated using artificial intelligence. You can make your own as fast as you can hit the mouse button by going to [thispersondoesnotexist.com](http://thispersondoesnotexist.com) and it will make up fake faces all day long, relatively easy. Constructing an infinite number of artificial personalities and histories is an easy exercise for a transformer as you would probably intuit from what we've said about them on the show. And the reason I said that it's not 100% important that this was about Trump was that while most of the disinformation is around right-wing content, QAnon, so forth, that it's been observed that Russia spreads disinformation on the other side as well. That they are principally interested in causing discord, argument. So they will want to feed both sides here, but to do it in a way that causes those sides more antagonism towards the other.

There is some hope. There was a study done at the RAND Institute that measured that people are less likely, on Facebook, to press the Like and Share button on content if they learn that it originated from Russia. There's a quote here from Todd Helmus, the study's lead author and a senior behavioral scientist: "Left- and right-wing audiences are particular targets of Russian propaganda efforts so they naturally have a big reaction because the propaganda speaks to them."

There's a lexicon here, a taxonomy of terms. *Disinformation* is false information knowingly spread. *Misinformation* is false information unwittingly spread. And *propaganda* is true information slanted to achieve a particular goal.

So, for instance, there was something going around the Internet some months ago that said, "This is the circuit diagram that was leaked for the 5G chip that Bill Gates is putting in vaccines to track everyone." First of all, I've got to say that Bill Gates does not want to know where you are. And if Microsoft wanted to know where you were, there would be a lot easier ways of doing it. But that aside, it didn't take long for people to research that circuit diagram and find out that it was actually a preamp for a guitar, I think. They tracked it down exactly. Well, the people that were spreading that picture throughout the internet were participating in misinformation. They received it from someone that they felt was trustworthy; they forwarded it without any kind of checking. But someone came up with that in the first place. The first person to put that out knew what it was, or at least they certainly knew that it wasn't anything that was coming from Bill Gates. And that was disinformation.

I know it sounds as though disinformation is worse than misinformation but actually, people who believe that what they are spreading is true, are probably more dangerous. If someone's getting paid to spread the rumor that COVID vaccines are fatal, that they cause deaths, then when that activity is made unprofitable for them, or too dangerous, they'll give up. But if someone *believes* that, they will fight to the death for it because those are the stakes. And that is not an exaggeration because we've seen lots of people dying through not getting vaccinated, and even on their deathbed believing that they didn't have COVID because that was a conspiracy.

And if you could radicalize people to the point of being willing to give their lives in the service of your idea, what a potent weapon. And talk about the United States military forces being ready to fight the last war, again. Nuclear weapons, intercontinental ballistic missiles, cruise missiles; none of this is any good against someone that can deploy a tool that turns your own citizens into them. At some point, they'll just change the flag at the top of the pole, and no one will notice the difference. Clearly, that kind of psychological warfare when assisted by artificial intelligence is cheap, easy to do, with little to no consequences if you're found out. And our responses right now are primitive at best. At least they're better than they were a year ago. We need ways of assessing people's mental states so that we can say, or they can find out, "You may have been subverted." But we don't have good tools for that. Our state of measuring mental health is at the moment like going to the doctor and saying, "Hey, Doctor, my heart

feels a bit down lately.” And the doctor says, “Oh, well we have these pills for people whose hearts feel down. They may cause weight gain and sexual dysfunction and the occasional suicide but on balance, they’re worth it.” We wouldn’t sit still for that; but that’s the case when it comes to mental health. That, by the way, was riffing on something I heard in a Lex Fridman podcast.

And then Facebook tags things as being about COVID-19. How much of a speed bump is that really? All you have to do is click the OK button. It’s like those warnings in front of TV programs that say, “Warning. The following show contains extreme violence, foul language, and full-frontal nudity. Viewer discretion is advised” in the most dramatic and enticing voice. How many teenagers listen to that and go, “Oh, okay, there’s the seal of approval. Got to watch this one”?

This is a war. Twitter automatically removes 10 accounts every second because they are bots. How many do you think they’ve missed? And what are those bots there for? Do you think they’re all there to conduct guerilla marketing for personal massagers and craft beer? Improbable.

Okay, so this has all been a big downer. And that roughly represents my state of thinking about it, so it’s not surprising. But what can we do to salvage some kind of hope about this? Well, technology can always be used to cure, or address, the problems it’s created. Mark Zuckerberg told Congress in 2018, “Over the long term, building AI tools is going to be the scalable way to identify and root out most of this harmful content.” Now, he’s not what I would call the most unbiased source of information about this because the amount of money that Facebook stands to make from creating those AIs is obviously substantial. But nevertheless, I agree with him. I encourage you to look for the Global Disinformation Index started by Alexandra Mousavizadeh who has created a continuously updated index of where misinformation and disinformation is found on social media and what advertisements are found next to it so that the advertisers’ attention can be brought to realize that their ads are running next to content that they may not approve of.

And we are getting more awareness; that counts more than anything else right now. There was a report in 2020 from researchers at Oxford University who tracked the number of countries that have experienced coordinated social media manipulation campaigns. It was 28 in 2017, 48 the next year, and by 2020, it was up to 70. Facebook was the chief platform for those disinformation campaigns. And there were seven states that were using that platform and others to spread disinformation. And here are those countries: China, India, Iran, Pakistan, Russia, Saudi Arabia, Venezuela. Russia, we’ve talked about. Let’s talk about China.

Very early on in the show, we spoke with Audrey Tang, the information minister - information, not misinformation, not disinformation, but information - Minister for Taiwan. And China would very much like to do to Taiwan, what it did to Tibet, and has done to the Uyghurs and is doing to Hong Kong. So Audrey has a tremendous amount of experience of foreign states, or a foreign state in this case, running disinformation campaigns against her country’s citizens. And when

COVID broke out, Taiwan got the usual share of conspiracy theories and other rumors that were not true, but they dealt with it much better than the United States did. And she outlines this in our podcast episodes two and three. The strategy is called “fight rumor with humor”, by making jokes.

So for instance, when there was a rumor going around that toilet paper was in short supply, and of course that would cause hoarding and other effects, they responded with a statement from their leader, saying, “We each have only one pair of buttocks,” and showing them wiggling their bottom, which... granted, I don’t think would work in the United States. I don’t think they should try. But it was very effective in Taiwan. But the other thing about that strategy is that she said they have to launch that response within 20 minutes. Any longer than that, and it’s not effective. Now, try and imagine that level of response from government in the United States and I think we have to think about other solutions. But even if we are forced to other solutions, at least the experience of Taiwan provides actual data on what works.

We need far more attention to our psychology and our state of mind. We are witnessing the effects of our computer technology advancing exponentially faster than our understanding of our own psychology. And that lack of understanding has caused us to leave vast numbers of people desperate, lonely, confused, and depressed, not either recognizing that that’s happened or knowing what to do about it. The author Maciej Musial in the book, *Enchanting Robots: Intimacy, Magic, and Technology* said, “As for the level of interactions with the world, it is well known that a significant number of sociological investigations of contemporary Western culture find that we are living in a society of risk, that our life is liquid, that we experience a lack of ontological security and increasing precariousness.” So technological response aside, addressing our mental and emotional health and just caring for people, more than trying to marginalize them, alienate them, or flat out eliminate them because we don’t know what it’s like to walk a mile in their shoes, would go a long way towards making disinformation campaigns less likely to succeed.

Don’t think of the people who are swayed by disinformation and conspiracy theories as being all stupid and ignorant. In many cases, this is leveraging a psychological need that they have, or amplifying a belief that they have about themselves. They’re drawn in by the prospect of being in the know, of being one of an elite group of people who are smart enough to take the red pill and see the truth that eludes everyone else. And so they quite likely think of themselves as being the critical thinkers, and the scientists and authorities as being the ones that are ignorant tools of a sleeping society. I haven’t seen any sign yet that we know what to do to turn that kind of emotional subversion around at scale. Honestly, it feels like *Invasion of the Body Snatchers*; you never know who next is going to suddenly turn into [one and] has completely alien logic and values. But I do think it’s time for the world’s psychologists, therapists, psychiatrists, to stand up, roll up their sleeves and say, “I’m going to do something about this. I’m going to help. I have the knowledge and the tools to do something about this.” We just need a way to put

them together to be of the right service. In the meantime, doing it one person at a time might be the best we've got.

We need to place more trust [in] and empower our authorities and experts. Now granted for the last 40 years, they have not done a lot to deserve it. They abused a lot of that faith that was placed in them during World War II and we can see what happened to institutions like the police and our governments. So there's movement that has to happen on both sides. But it won't be complete until we put our trust in institutions and science. I say this because I was about to say, "Go and do your research," and then I realized that that is one of the slogans of the QAnon movement which uses it as a catchphrase to enroll people in their own conspiracy theories. Because if you research on YouTube, the search terms that lead in that direction, they will take you all the way down that rabbit hole. So research is not something that happens on YouTube. Research is something that scientists do and then they write papers. And part of the problem has been that most people couldn't understand those papers. So it was left up to a few people to translate those into words that the rest of us could understand.

And part of the problem has been that people who don't understand the scientific process will listen to those explanations and think that that's how science is done - by coming up with a plausible explanation. So for instance, when someone hears an explanation of quantum entanglement that's been simplified for the layperson, they think, "Well, that makes sense. The universe, we're all connected, right? That's what it means. I knew that." But because they didn't see the papers and the equations and the research and the experiments that went into confirming all of that, they think that science is the result of coming up with a plausible explanation, which is what we did prior to The Enlightenment. But then, clearly, anyone can come up with an explanation as long as it sounds plausible, then you're done. Which is why people have so much trouble deciding what's true or false, and finding that they might as well believe all of the conspiracy theories because they all sound plausible. We won't be out of this until we understand that scientists and other experts who've dedicated their lives to a process of uncovering the truth through the scientific method, are, by Occam's razor principle, the ones to trust, even and especially when we don't think they're right, because, by Occam's razor, they're more likely to be right than we are.

Anyway, if you're still with me after that prodigious amount of ranting and rambling, thank you. You know, some AI researchers have said that the reason they love researching AI is that the more they learn about artificial intelligence, the more they learn about themselves. And I hope that's true about people who are researching AI used in disinformation and misinformation remediation.

I suspect this is going to get worse before it gets better. I believe that it is the root of many, if not all of the strongest, hardest problems that we have today. Look at climate change. Understand that if the 2016 election had gone the other way, the United States would not have pulled out of the Paris Climate Accords for one thing. And then look at how much misinformation there is about climate change. And what is that doing to our efforts and

coordination to do something about the problem? It's not hard to see disinformation as itself being an existential threat. Now you know why I end every episode by saying, "No matter how much computers learn how to do, remember, it's how we come together as humans that matters." Because - forget *Terminator* scenarios of AI destroying us; this may be the way that AI destroys us - as a tool used to cause us to set at each other's throats.

So if you've got any skills, any abilities whatsoever to address this problem, this is your wake-up call. This is not a theoretical problem years or decades in the future. This is happening now. Now is your time to stand up and get involved.

Whew, okay. That was pretty heavy. I hope the next time I come back to this I have something more positive to report on this. The episode next week will be with guest Jonathan Rowson, who is a chess-grandmaster-turned-philosopher-activist. And I will get to talk to him about both of those things which intersect with artificial intelligence. Because we'll talk a lot about how computers in chess have changed our approach to chess, games, life, and how we can coexist with machines that are coming to think more and more like us.

In the meantime, and especially if you've stuck with me through this whole episode, this is for you: Remember, no matter how much computers learn how to do, it's how we come together as *humans* that matters.

<http://aiandyou.net>