Guillaume Chaslot: It's like you have this huge current that pushes you towards being more aggressive, more divisive, more polarized.

Aza Raskin: That's Guillaume Chaslot, a former software engineer at YouTube. If you've ever wondered how YouTube got so good at predicting exactly what will keep you around, ask Guillaume. He worked on the site recommendation AI, and he marveled that it's power to sweep a viewer along from one video to the next, setting them a drift on a stream of idol viewing time. He celebrated as the streams multiplied and gathered strength, but he also detected an alarming undercurrent.

Guillaume: It was always giving you the same kind of content that you've already watched. It couldn't get away from that. So you couldn't discover new things, you couldn't expand your brain, you couldn't see other point of views. You were only going to go down a rabbit hole by design.

Tristan Harris: To understand where these algorithms might take a viewer consider for a moment how they're designed. Think of that moment when you're about to hit play on YouTube video and you think, I'm just going to watch this one and then I'm out and that'll be it. When you hit play inside of YouTube's server, it wakes up this avatar voodoo doll version of you. Based on all your click patterns and everyone else's click patterns that are kind of like the nail filings and hair clippings and everyone else's nail filings and hair clippings.

Tristan: So this voodoo doll starts to look and act just like you. And then they test like they throwing all these little video darts at you and see, if I test these hundred million darts, which video is most likely to keep you here? So now in this case, YouTube isn't trying to harm you when it out competes your self control. In fact, it's trying to meet you at the perfect thing that would keep you here next. That doesn't have to be bad by the way, right? They could just show you entertaining things and so suddenly they've taken control of your free will, but it's not bad because they're just showing you cat videos or whatever.

Aza: Guillaume observed a subtle but unmistakable tilt in the recommendations. It seemed to favor extreme content. No matter where you start, YouTube always seem to want to send you somewhere a little bit more crazy. What Guillaume was seeing was algorithmic extremism.

Guillaume: When I saw that, I thought, okay, this is clearly wrong. This is going to bring you many humanity to a bad, bad place.

Tristan: Now this is exactly what you would hope to hear from a conscientious programmer in Silicon Valley. Particularly, when that programmer is building an algorithm that can determine what we watched to the tune of 700 million hours a day. Guillaume could see how these crosscurrents would pull viewers in countless delusional directions. He knew the algorithm had to change and he was confident he could change it.

Guillaume: So I proposed different type of algorithms and a lot of Google engineers were motivated by that. Like seven different engineers helped me for at least a week on these various projects.

Aza: You'd hope this would mark the beginning of a humane design movement at YouTube's headquarters. So what happened?

Guillaume: But each time it was the same response from the management, like it's not the focus, we just care about watch time. So we don't really care about trying new things.

Aza: Anticlimactic, right? And that's actually how this conversation plays out. Or fizzles out. Again and again. In Silicon Valley, managers rarely ever reject solutions outright. They just do what managers do: set the team's priorities. A slow creeping tilt towards user extremism, that'll never be on the top of this quarter's priority list.

Tristan: Today on the show, we'll ask Guillaume to game out the consequences. I use the word game deliberately, because in the sense YouTube recommendation system is engaged in a chess match for your attention. It's trying to predict your next move to catch the moment you've had enough and are ready to leave the site and to overwhelm your self-control with the next irresistible video. And you may think, "Big deal that's on me." Well, take it from the designer of these algorithms, you're up against the machine that can outsmart a chess master.

Aza: So today on the show, Guillaume Chaslot, AI expert and founder of the nonprofit AlgoTransparency will explain why, for the sake of humanity, we must shed light on these algorithms, understand their inner workings, and more importantly make visible their outcomes. So we can tilt the rules of play back in our favor.

Tristan: I'm Tristan Harris.

Aza: And I'm Aza Raskin. This is Your Undivided Attention.

Tristan: Why are they even doing these recommendations? I mean, you could imagine landing on a video site, you watch a video, but there's no recommendation. So why is recommendation so important to YouTube?

Guillaume: So more than 70% of their views come from the recommendation. That's huge. Knowing that they do 1 billion hours of watch time every single day. 70% of that is like a tremendous amount of watch time.

Aza: Yeah. There's sort of like this is the content that we are 700 million hours of dosing humanity with something that humanity hasn't chosen.

Guillaume: Exactly. So you have very little choice on this content because ... So YouTube algorithm has 10 billion videos or I don't know how many billion videos only chooses the 10 to show to you in front of your screen and then you have just a tiny little choice between those 10 to choose which one you want to see. So it does 99.99999% of the choice is from an algorithm that you don't understand or you don't control.

Aza: So ok, I'm just getting shown stuff that I like clearly have a revealed preference for that I'm clicking on or watching or that works on other people besides filter bubbles. What harm does that create?

Guillaume: So it creates a bunch of harms like when I show conspiracy theories, because conspiracy theories are really easy to make. You can just make your own conspiracy theories in like one hour, show it and then you can get millions of views. They're addictive because people who live in this filter bubble of conspiracy theories don't watch the classical media, so they spend more time on YouTube. So every single one of their watch will have more total watch time. So it will have much more weight on the algorithm. So the more people watch them, the more they get recommended. It's like a vicious cycle.

Aza: So you're saying that conspiracy theories are very effective at grabbing our attention and keeping us around and they become kind of like black holes that if the system's just recommending the stuff that people click on, one of the techniques is going to find is recommend conspiracy videos because conspiracy videos are very effective. Is that what you're saying?

Guillaume: Exactly. That's the same way a black hole creates and the only grows bigger, like I design this conspiracy theory can only grow bigger because then people who are in there spend more time than others. Imagine you say you're someone who doesn't trust the media. You're going to spend more time on YouTube. So since you spend more time on YouTube as the algorithm think you're better than anybody else for the algorithm, that's the definition of better for it—who spends more time. So it will recommend you more. So there's this vicious cycle. So it's not only like don't trust the media, but it's with any moral, the algorithm by design will be anti moral.

Guillaume: So if you have like a moral in the society says like racism is bad, humans are equal and people think that. Racism is good, they will spend more time on YouTube so that we get recommended more by the algorithm. So like the anti moral will be favored by the algorithm. So Google is saying, yeah, we give a place for these people who are not accepted by society. We give them a place to express themselves on. That's, I have no problem with that. But what I have a problem is that it's structurally, systematically anti moral. So even if we reach a numeral, let's say we go towards a moral in which like, okay, racism is great, then the anti moral will win again. It's just ridiculous.

Aza: But I think I'm hearing you saying is that because the AI doesn't have a sense—the recommendations—it doesn't have a sense of what's right and what's wrong. All it has is sense for is what works?

Guillaume: Yes.

Aza: That we're sort of A/B Testing our way with the smartest supercomputers pointed at our minds to find sort of the soft underbellies to just be like, what's effective? And so our A/B testing our way towards anti morality or immorality or amorality.

Guillaume: Yes.

Aza: Are there any specific examples that can light up my mind?

Guillaume: Like the flat earth conspiracy theory for instance, got hundreds of millions of recommendations for something-

Tristan: Hundreds of millions of recommendations.

Guillaume: Yes like for something that's completely absurd. So one of the arguments was like, we're just showing what people make, but that's not true because if you search on YouTube flat on earth you had 35% of search results were flat eearth conspiracy theories. But then if you followed recommendations like I followed thousands of recommendation and then took like the 20 most recommended videos and then out of these 20 most recommended video 90% were flat earth conspiracy theories.

Tristan: 90% that's insane.

Tristan: Well, so I think one thing that people tend to think about with this is ... I mean if you just go back to the just the simple human experience of YouTube, like why are we spending all this time and the average watch time per day is 60 minutes. Now the YouTube product officer, a chief product officer, Neal Mohan said, it's because of our recommendations we're getting that good. So, the reason that watch time is going up is because the recommendation system is getting stronger and stronger and stronger every year. And we're not talking about the fact that this isn't huge asymmetry of power. They have supercomputers, I mean who has the biggest supercomputers in the world? It's Google and it's Facebook.

Guillaume: So we blame ourselves, we blame teenagers. We are like, hey, this-

Tristan: You should have yourself control.

Guillaume: Yeah. You have bad parenting. You are bad person, but you have a supercomputer playing against your brain as you said. And it will find your weaknesses. It's already studied the weaknesses of billions of people, it will find them. So my weakness for instance, is a plane landing videos. I don't know why. I'm fascinated by plane landing videos. There's lot of that on YouTube and if you would ask me, "Do you want to watch accidents on plane landing videos?" I would say no, never show me that, I don't want to waste my time watching that. But you can't say that to YouTube, so it will show it again and again and I lost so much time watching this plane landing video. This is ridiculous.

Tristan: So YouTube is discovering these weaknesses for so many different demographics, right? And so you have this example of teen girls who started watching dieting videos, like what kind of food should I eat? They get recommended anorexia videos because they're better at holding onto that demographic and they recommended this millions and millions of times. You have this other example, you know of you watch a 9-11 news video and it recommended 9-11 conspiracy theories and the number with Alex Jones for example, always stun me. You said that it recommended Alex Jones videos, 15 billion times.

Guillaume: Yeah. And that's a lower estimate. I think it's much more, but we have no idea how big it is.

Aza: Why do we not have any idea?

Guillaume: Because YouTube doesn't want to say how many times they recommend each video. So, there like, yeah, if we start saying it, then we give way too much information and they have no incentive to actually do it.

Aza: What are they afraid of?

Guillaume: For the small recommendations they might be afraid of people gaming the algorithm, but people are already gaming the algorithms. If you search on Google, should I buy YouTube views? You have an information panel that says, yes, you should buy YouTube views. It's like Google's algorithm says that you should game YouTube algorithm.

Tristan: One of the things I also find fascinating about your research is when you look at YouTube incentives, it's not just that it wants your watch time as a viewer, it's also trying to give YouTube creators that hit of dopamine because when you host, when you publish a video, it needs to give you that like rush of feedback of look at how many people are starting to watch your videos. So it gets everyone addicted to getting attention from other people. Guillaume, if you can explain the cold start problem that YouTube is trying to solve.

Guillaume: So when you put a video online, it has no views or very little views. Maybe you send it to your friends. It's like five views, but from five views it's mathematically impossible to detect how good the video is from just these five views. So the algorithm has two ways to be behave. It could either really try to be fair and give an equal chance to every single video. I'll say, okay, it only had five views on it. It doesn't seem so good for now, but we are going to show it to a lot of people to see how well it performs. So that will be like the fair way of behaving for YouTube.

Guillaume: But this fair way of behaving is way too costly because by doing that, it means that you promote really bad videos millions of times. Like if you had all the bad videos that you have on YouTube, you have to promote them each, let's say a thousand times before you can have good statistics. So if you do that on thousands of videos, it's like millions of times, it's probably millions of videos. So you do that billions of times. So you lose a lot of money and a lot of people get away from YouTube. So that's not what YouTube is doing. YouTube's Algorithm is pretty greedy. So if you don't-

Tristan: It's a greedy algorithm.

Guillaume: Yeah, so that's the scientific term. It's actually greedy. So it means that if it doesn't have extremely good stats from the start, the algorithm is going to stop recommending your video like right away. So you got this, the very few first views are very crucial. That's why Google's algorithm say, you should buy YouTube fake user views because that's the best way-

Tristan: That's the way of juicing your cold start.

Guillaume: To juicing the start of your video.

Tristan: I mean what, I always find sad about this is when you just look at what it's done to culture, because we have this point that in the race to get attention, it's not just enough

to get your attention. We have to get you addicted to getting attention to other people. And so it turns society inside out where we all want to get as many views, followers, subscribers as possible. And because YouTube with this cold start problem, it wants to make you feel like, hey, don't you feel famous? Like we get you a thousand views and you want that number to go up. And we're all so vulnerable to that. I mean, we're just like chimpanzees getting this number to go up and like, oh, that feels so good. I want to like refresh it again and see how many, if I got some more views now and listen by the way, I've studied this for like 10 years.

Tristan: I do this still, even last night we had this big Wall Street Journal piece on our work. And I went back and I checked again to see how many people had been looking at it and what kind of feedback it had gotten because we really care what other people think about things. And that was a design choice YouTube made because they wanted to get people addicted to how many views that they had. And I don't mean nefariously like bad evil people, but unless you want the whole system to be kind of operating in a way where you automate all the chimpanzees to just like want to put up their videos and themselves and the makeup videos and all these things. And it really turned culture inside out.

Aza: Hey, this is Aza. So we're going to pause Guillaume's interview for a second. Tristan testified before the U.S. Senate on persuasive technology last week and I wanted to ask him to tell us about it. I've never gone up and given a Senate testimony. What is that like?

Tristan: I honestly was really impressed with how especially a few of the senators had really understood it. I mean, I felt like Senator Schatz and Senator Thune who are the two chairs of the Commerce Committee that I testified at knew the topics very, very well. Senator Blumenthal knew the topics very well, Senator Markey knew the topics very well and I think that this helps displaced this notion that government doesn't get it. Now yes, are there some people on that committee who knew far less or may have missed or made gaps in their comments? Absolutely, but I especially, I think Senator Schatz, he's from Hawaii, and he gets the entire thing.

Senator Schatz: Social Media and other Internet platforms make their money by keeping users engaged and so they've hired the greatest engineering and tech minds to get users to stay longer inside of their apps and on their websites. They've discovered that one way to keep us all hooked is to use algorithms that feed us a constant stream of increasingly more extreme and inflammatory content. And this content is pushed out with very little transparency or oversight by humans.

Tristan: I was talking recently to the former FCC chairman, Michael Powell and we were talking about standards and practices for children's television. I believe still the case that Nickelodeon or a Children's TV network is not allowed to put a URL inside of a TV program for danger that you might be pushing a child towards a website that you don't know where it's going. Can you imagine that compared to YouTube, where YouTube is push, push. It's all it does is it pushes you and shoves you and even worse than that, it has all those buttons and links that come up saying, subscribe here, click on this. Do you want more child pedophilia? Do you want to know even more how to commit suicide? It is a war zone. We would have never, you know, in my Senate testimony, I made the analogy to Mr. Rogers. I mean, I just love this example because he came to the exact

same committee, the Commerce Committee 50 years ago and he said, he was so concerned about what he called the animated bombardment that we were throwing in front of children.

Mr. Rogers: We deal with such things as the inner drama of childhood. We don't have to bop somebody over the head to make drama on the screen. We deal with such things as getting a haircut or the feelings about brothers and sisters and the kind of anger that arises in simple family situations and we speak to it constructively.

Senator Pastore:How long a program is it?

Mr. Rogers: It's a half hour every day.

Tristan: He convinced the hearing, the committee. In six minutes he takes the most cynical senator who's ready to defund all of PBS and he just says at the end, "Well, I guess you got your $20 million." The comparison is mind-blowing. By contrast, we are sending children into a kind of a war zone of unpredictable, mindless and extreme stuff. I just want to add this other note from my conversation with the former FCC commissioner, that if you looked at the time spacing between commercial breaks and television shows 30 years ago, the screen would go black before the commercial break and there'd be like a pause, a real pause where there was just nothing there. And that originally as I understand it, was in part related to the way that in theaters the curtains would drop and there was a break and you have to sort of, you go to the theater and you come back, the intermission, et cetera.

Aza: You mean you have to decide?

Tristan: Yeah. It forces you to make a conscious choice and especially when it comes to children in front of a television, that break is critical. I mean-

Aza: So, you're saying there used to be breaks in between shows?

Tristan: Yeah.

Aza: Okay.

Tristan: Yeah, exactly. So, these are the kind of stopping cues that we've deliberately lost per our episode with Natasha and the design towards removing right angles and all purposeful breaks and it just really struck me that, wow, this is something that we really need as humans.

Aza: Now back to the interview with Guillaume.

Tristan: One of the other things I find fascinating is a point Aza actually has made that the attention economy grooms humans to be better for the attention economy. Like, there's almost two ways to predict a human's behavior. One is, given this human, let's build a more and more accurate model so that we can predict better and better with bigger computing power and more data, meaning more nail filings, more hair clippings to

make them look more and more like you. We're going to build a bigger supercomputer so that whatever we couldn't predict yesterday, we can predict more of today.

Tristan: But, the other way to better predict people, is to make them simpler and more predictable. So, if you make them act out of their fear, out of their amygdala, out of their dopamine sort of wriggling around inside their nervous system, they're more reactionary impulses, they're also more predictable. So, we're being groomed on a certain level to be more reactionary, more outraged, more concerned with our status and how we are perceived by other people, more addicted to getting attention and status.

Tristan: We might want to talk about this thing, how is YouTube actually responding to all this? Because if we're tilting the landscape, they've hired 10,000 content moderators, if not more now, probably mostly in English. The joke about this is that, these are like hiring 10,000 boulder catchers so that while the landscape's been tilted and all the outrage ann polarization and crazy conspiracies are flowing downstream. They hired 10,000 people to catch the boulders and it's like they're not going to catch nearly enough of them.

Guillaume: No. They catch only the very extremely visible things, that's a danger that they remove all the thing that's visible and then you don't see like all little boulders that go down. For instance, we saw like one things they didn't catch for like the last 10 years, like this February was this pedophilia problem. So, basically the algorithm was recommending little girl videos to people who were watching fitness videos because there was one chance out of, I don't know, maybe one out of 100 that you would become a pedophile or you would be a pedophile. You would like watch these videos on pedophile, watch videos for a long time, they watch little girls for a long time so, the algorithm was actually recommending little girls to many people. Until this one YouTuber discovered that nobody was talking about it and YouTube reacted, advertisers reacted. It was a huge deal but it's like one more thing.

Tristan: I mean this speaks to something critical structurally about this problem, which is per your point. When does YouTube respond to these problems? It's usually because someone like you stays up till three in the morning in an unpaid, nonprofit civil society researcher, who's just staying up saying, hey look, I think I'm seeing some problems. I'm building my own tools, I'm not paid by a company, I'm just doing this because I care and I understand some of this. There's a handful of people like you, like Renee DiResta, people scanning for what is Russia doing, what is China doing, is Iran doing or what are the recommendation systems doing. They're unpaid and they find these things, they have to work hard to get the Washington Post or New York Times to report on this stuff or it works hard to hold a hearing or get Senator Mark Warner or Adam Schiff to write a letter to get the companies to respond. They work so hard to do this, but they're only catching a handful of these issues and mostly in English and in Western markets.

Tristan: But now you consider, that YouTube is actually the most popular thing in Mexico and in the Philippines, the watch time is like off the charts. It's the most efficient medium and how many people like Guillaume or like Renee, these people who are doing the hard work exist in these other markets. So yes, we have maybe the best boulder catchers in the world here in the United States, but we only have a few of them and we have none of them in some of these most vulnerable countries. Where we know the polarization is most extreme whether that's Myanmar and the genocide happening there with the

Rohingya minority group or the way that in Christchurch with the New Zealand bombings. I mean, they've created a digital Frankenstein that they cannot control.

Guillaume: Yes. To give an example of that, this asymmetry like measles outbreaks like rose-

Aza: Measles outbreaks. Got it. Yeah.

Guillaume: Measles outbreaks rose at 300% in the first trimester of 2019 globally-

Tristan: 300% more measles cases. Wow.

Guillaume: Yes. In some parts of Africa, it's 700%-

Tristan: 700%.

Guillaume: So it's huge. So you see the asymmetry between countries that have like structures on good media that can fight these social media problems, but in Africa doesn't have good political social and good press. Then they just can't fight it and then people really sincerely believe that vaccines are designed to kill you.

Aza: So yeah. So quickly just draw me the line from like YouTube recommendations to these incredible stats?

Guillaume: I noticed like from two years ago that YouTube recommendations were showing a lot of anti-vaccine conspiracy theories. So there were different types of anti-vaccine, they were like, for instance, Bill Maher who was saying, "hey, don't take the flu shot, it's like, it's a bit of a kill, I wouldn't take the flu shot." I'm like, okay, fine, but there is really, really dumb video saying that vaccines are designed to kill you or look at my little child before and after autism. You have like this very emotional video, with this very emotional music of a little girl who was beautiful before having autism and then started to have autism on that generated. The parents said it was because of the vaccine but there is absolutely no scientific evidence there.

Tristan: Now if we take this more full stack, sort of socio, emotional understanding of why this is happening and think about in a parent? Why is it so compelling to watch a video like that? Because the idea that you would inject your child with something that would give them autism is so fear inducing.

Guillaume: Exactly.

Tristan: That if there's even the tiniest, tiniest chance that that can happen. If that surrounds you and if when you, as you said like you started on a YouTube page, right? For Bill Maher's saying, "hey, don't do the flu shot." That's a reasonable video to watch and I'm sure he said it in a funny way. But then, for someone who watches that video, YouTube calculates, well what would be the really good videos to show someone who saw that Bill Maher flu shot video and it discovers that these are incredible, emotional, powerful videos of parents and it's preying on fear, it's preying on emotion, what if I screw up? It's the loss.

Tristan: I mean it's horrific. So, it's totally understandable why A, these videos would be recommended and B, why parents are going to react so strongly to this? It's a very powerful stimulus to throw in front of the human nervous system. This is also people need to understand is how, what a global problem this is? You have this example of the Syrian refugees and I think some Russian conspiracies about Syrian refugees. The white helmets in Syria are a peacekeeping force?

Guillaume: Yes, exactly.

Tristan: But if you actually Google for or if your searching YouTube or looking at those videos, YouTube recommends what?

Guillaume: Yeah. So, the Russian propaganda outlet made a very good case for the White Helmet being like this, these terrorist force that are like secretly helping terrorists to do terrorist things. So they made so many of these viral videos that were without foundations behind them, if you start looking for White Helmet, YouTube will tell you that they're a terrorist group. I met someone from the White Helmet, she told me she had member of her family telling her like, hey, what are you doing? Are you terrorist? So, the people start to believe more-

Tristan: What they see on YouTube.

Guillaume: What they see on YouTube than their own family. Imagine that.

Aza: Yeah. There's one stat that you mentioned to me yesterday about the Mueller Report coming out and which channels were most recommended?

Guillaume: Yeah. So, basically a swing in the data that Russia today video that didn't get that many views would get around 50,000 views. It got recommended for more than 236 different channels. It was more than any of the 84,000 videos that I was monitoring.

Tristan: So, this is the Russia today video was more recommended than any other video about the topic of the Mueller Report.

Guillaume: Exactly. So, Mueller Report is about the Russian interference so, YouTube is recommending the take of the Russians on their interference into the 2006 in action, that's pretty ironic if it wasn't so important.

Tristan: So, I think there's this sort of automated machine, we use this rhetoric that these systems are putting thoughts in people's minds some people say, oh no, you know, it's hijacking our mind, people say, oh, you're just over exaggerating like what are you talking about? Like I'm choosing my own thoughts, I'm living my own life, maybe I don't even use YouTube, but when you really realize the scale of this and that the actual reality that the evidence that people's subtly psychologically influenced on the emotional level and the physical level and behavioral level to believe bits and pieces of this.

Tristan: Even as you said, if one 1 of a 1000 people believe the Alex Jones things and just imagine if a 1000 people and only one of them believes the Alex Jones, that's 15 billion divided by a 1000, that is such an insane number. I mean, your brain literally talk about humane

technology, our brains are not tuned to deal with large numbers. We cannot even reason about them. We have such poor reasoning about large systems. So one of the hardest things for me about this problem is how do we get people to see the vast scale and the influence, at a global stage level because people want to focus on their own experience, like I'm not addicted to YouTube and we're trying to get them to say it's not about you, it's about this much bigger system. What do you say to that?

Guillaume: Yeah. So first thing that, if you listen to this podcast, you're not one of the most vulnerable people and you can say it's okay, it's just a vulnerable people are getting tricked by the algorithm, but the algorithm is improving so fast that soon it will be you. So, that's why you need to pay attention to the vulnerable people right now. That's why we need to pay attention to anti-vax, that's why we need to pay attention to conspiracy theories. People don't realize how different it could be. We realize it because we were at Google, we were there when the choice were made and we realize our different the choice could have been, but people think it has to be that way because that's-

Tristan: All they've ever seen.

Tristan: Well let's talk about that for just one second because Guillaume, you and I both shared this experience of seeing some of these problems while being inside of these companies. Right? A lot of people think Guillaume probably, when they hear your or my story, they think, oh it's these greedy companies and they just want this, they just want their money and that's why they're not changing. But my experience, I would talk to say someone who ran android and I would say, okay, you are in control, whether you handed the puppet strings of 2 billion people to these apps, you are the government regulator of the puppet strings. You get to decide which strings you are they're allowed to pull and which ones they're not allowed to pull and what are you going to do? I would explain this problem and people would look at me and they would nod and they'd say, yeah, that's a problem that's an issue.

Tristan: And then they'd say," I'm really glad you're thinking about that" and then we would make these proposals of here's how android could be different, here's some notifications rules, but nothing ever got implemented. It wasn't because in my case someone said, "that's going to drop revenue." It was mostly, "oh, is that really a big problem?" We've got these OEM suppliers, we've got these new phones to ship, we've got the next version of android to ship next year. It just never became a priority and I'm curious, in your case, there you were in YouTube and you're starting to raise these issues. How did people respond and what did you try to do on the inside?

Guillaume: Exactly. So from the inside I tried to be a very positive, mostly because we have this image as a French are always complaining about us. So, I didn't want to be this typical French that complained about things.

Tristan: I read some conspiracy theories that the French are the most complaining-.

Guillaume: No they are. It's true. So, I didn't want to be like that, I wanted to be like positive into solutions. There was this saying at Google, there was this thing, I love this, if you see a problem, fix it. If you see a problem, don't complain about it to the management, just fix it by yourself and that was supposed to be rewarded. That's what I did, I saw this problem and I propose an implemented solution with some people and I thought that

was going to work, but then as you said, people said, yeah, is it really a priority we're trying to make the product grow like 30% a year? That's huge-

Tristan: Watch time.

Guillaume: That's a priority like watch time grow 30% a year, that's fantastic. This is like a distraction for me, for them it was distraction like trying to help people get out of their filter bubbles.

Aza: Right. So how do we go about creating protections or regulations around recommendation systems?

Guillaume: Yes, so that's very tricky because we don't want to block free speech of course and that should be the absolute priority. But recommendations are not free speech they are free. There's a freedom of Google to make money, with anti-vaccine content. So it's not a problem of free speech to regulate recommendation, it's problem of free speech to regulate what can be put on the platforms and that's why CDA 230 that regulates platforms right now is actually really good positive legislation. But at the time when it was voted, recommendations didn't exist, AI didn't exist, so it was a very different thing.

Tristan: Lets explain what CDA 230 is. So this is the Communications Decency Act of 1996, and section 230.

Tristan: When someone says CDA 230 they just mean that. And this is specifically carving out a no responsibility, platforms are not responsible for the content that appears on them. This is what an allows the internet to grow.

Aza: Seems like a great thing.

Tristan: But as you said, it was before the age of AI. It was before anyone had built recommendation systems. It was before there was YouTube because 1996, is actually 10 years before YouTube.

Aza: And it's a completely different thing to be like, yes, platform you're not responsible for the user uploads, than saying platform you're not responsible for taking something you saw uploads and promoting it or amplifying it. This is where we come to that phrase of the freedom of speech is not the same thing as the freedom of reach. Imagine if we said cool, like what's true for the New York Times and other media is true for YouTube. Anytime that you as a platform make the curatorial decision, whether it's with humans or with an algorithm to amplify content, it's at that point that you become liable.

Guillaume: Yeah, exactly. So amplifying how many times do you become liable is a valid question. Everybody would agree if an algorithm amplifies something a billion time, it should be liable-

Tristan: It should be a publisher.

Guillaume: At some point the number of times that the algorithm say that Obama was born in Kenya, was in the hundreds of millions.

Tristan: Hundreds of millions.

Guillaume: So, it's completely crazy. It's probably more than the population of the U.S. It's insane. At some point if the algorithm is not liable, these things are going to happen. The idea is to have accountability, like at least, we can have an AI like in charge of where we're going, but at least we should know where it's heading. So we should know if when you constantly recommended on YouTube, we should know which proportion of the view comes from the recommendation and which proportion of the view come just from human recommendation to another. So there was this law passed in France that says exactly that. YouTube should say the proportion of each algorithm promoting the content for-

Tristan: So if you had a video that's got 100 million views, then you should be allowed to be able to say what percentage of those views came from recommendations.

Guillaume: That's right.

Tristan: So this will open up a bunch of transparency and accountability for YouTube.

Guillaume: Exactly which percentage of the view come from the search results et cetera. So you would have more visibility into what's going on and so if something goes wrong like, there's a bit of a case or bad child videos and stuff, you would see it much faster because you would see that the algorithm is starting to amplify like crazy, some specific type of videos. So you wouldn't need to wait until the problem is too big, you could see it faster.

Tristan: Another problem it seems is just speed, because if you think about the most profitable business model for YouTube it's to have all of this running on automation. So, you publish a video and it gets instantly available and recommended everywhere instantly as fast as possible with no human reviewers. That's the most efficient business model. Then you have no human beings in between guarding between what is being broadcasted and the sensitive people on the other end. And that includes children on YouTube for kids, that includes in Syria what people are believing about these sort of war zones where there's not much information coming out. So having a more sensitive, more protected way in which information gets controlled or shared or there's more thoughtfulness and not just an automated channel that's just trying to maximize profits. So this is why I think what you're doing in France is so critical and why we could replicate that in the U.S. or the EU.

Aza: And why I think AlgoTransparency is so interesting because it's essentially a citizen is having to create the satellite network to point back to understand what's going on earth. Like that's ridiculous.

Tristan: Yeah that's a perfect analogy I mean, so Guillaume has this project called AlgoTransparency, which basically shows as much as it can, it scrapes YouTube and it shows these are the things that are getting most recommended. And it tracks it over time so we can start to see trends. But this is one human being, one very talented human being, a civilian, who's trying to create essentially a system of satellites to monitor what's happening at the scale of 2 billion people. This is just not the right way that accountability should work.

Aza: And I know that when a YouTube will often fight back at you and be like, oh, but you don't have the latest data. And that's sort of your point. You're like-

Tristan: Correct. Yes.

Aza: You do.

Tristan: And they're trying to hide it.

Tristan: Hey, this is Tristan. What if we lived in a world where Exxon was the only company that knew how much pollution was actually out there because they owned all the satellites. That's actually kind of like YouTube and the pollution that is dumped into our public sphere. Let's talk to Aza about amplification transparency.

Aza: So Guillaume has done the most research on YouTube, but of course the same engagement bias, the enrage to engage is happening of course on Twitter and Facebook and all the techno social platforms. Just like there's an algorithmic bias against race and gender, there's an algorithmic bias against our values and every time our values are pitted against engagement our values loose. And here's I think the most important point to remember, it's whether platforms are choosing the content to amplify or choosing an algorithm which chooses the content to amplify. They are still choosing. I think that choice has greater impact than the impact of any major news organization and probably all of them put together or to put in other way, the platforms are choosing what goes into the information soil from which all of our collective sense and decision making abilities grow. So that's a lot of power, a lot of responsibility for which the platforms right now have no responsibility-

Tristan: Right.

Aza: So amplification transparency is an idea that we're interested in to do just that, to put back responsibility where it needs to be. And the idea, at least for me originally came from Guillaume. Algorithm transparency is about being able to start teasing apart the question of like why is the algorithm doing this versus that? And amplification transparency is saying, what is the algorithm doing? Just give me the hard numbers of which content is being promoted so that we as civil society can come together and decide, "hey, is that decision in our values or is it attacking our values in favor of engagement?" That's the difference. And the idea I think is very simple force platforms to expose an API where anyone in civil society can ask, hey, how much have you amplified recommended a piece of content, now and historically? And the platforms are required to answer, it's just an API that quantifies and makes visible the platforms bias so that we can decide together where their choices fit our values and where those choices attack our values.

Aza: It lets there be tens of thousands of Guillaumes. So that's oversight the scales to the scope of the problem. This is your analogy, Tristan, but if social media and tech platforms are a patient that has cancer, we want to save the patient by just removing the cancer. But in order to do that, we have to be able to see where the cancer is, else you just cut out the stuff that's helpful to living and you kill the patient. And the good news is for me that if we fix it, the results can happen fast. Here's an example, of how fast that can actually make change when you shut off the algorithmic hate and it's from that

countrywide study across Germany by Mueller and Schwartz at Princeton. And so I'm quoting from a summary from the New York Times, but whenever Internet access went down in an area with high Facebook use attacks on refugees drops significantly. And that gives me a lot of hope because that says, oh yeah, we are confusing mirrors for screens and if we just make that difference apparent, we revert to being our real actual selves.

Tristan: Right. I love that. Yeah. Replacing mirrors with seeing them as amplifiers, not mirrors.

Aza: Yeah.

Tristan: They're like fun house mirror, they're like the story.

Aza: Exactly. They're fun house mirrors. That's a good metaphor.

Tristan: Look by the way, YouTube is a fantastic and amazing product that provides life changing experiences. So just to say and make sure that we're all with you here in the tech industry, this is not an anti-tech conversation. It's about what is this automated attention, hungry AI powered system doing to history, to world culture. The problem occurs when they have a self dealing extractive business model that says instead of wanting just to help you with your goal, we really just want to suck you down the rabbit hole. And there's no reason why recommendations should be on by default. Like this is not a, you shouldn't be able to post ukulele videos or posts the health how to videos. This is about why are we recommending things to people that systematically tilt in the more extremizing directions that we know are ruining society.

Tristan: So, how do we actually regulate it? I mean why not, just not have the recommendations at all except when you click a button specifically. So the default setting is no recommendations, just like the default is to auto play and you can flip that off of course, almost no one does and YouTube knows that. And I say, oh, but we gave you a choice and it's your choice. It's just all manipulation. They need to be much better about this. Where is YouTube offering these lasting value, these lasting use cases? And how do we strengthen all of those cases where it's helping more people with their health injuries, helping more people learn musical instruments, helping more people laugh with friends. There are so many good use cases but it's not optimized for that at all.

Guillaume: So there are several classes of solution. There's one class of solution is to optimize for better things. So instead of optimizing for how much time you will watch, which we will lead to this like false medical information, you optimize for good feedback like yeah, this YouTube video really helps me and this should count a lot. Right now, It doesn't because YouTube doesn't care that it helps you, it just cares if video keeps you online longer. So, there was this video having 8 million recommendations that says don't drink table salt because there is glass in it and it's hurting your intestines. This was very scary so it worked really well and it got 8 million views or something like that.

Tristan: And the problem is whether or not human authority is a good rating system. I mean, you've all Yuval Harari, author of Sapiens and I talk about this, that this is a question about the breakdown of human choice and feelings. Because if enough people can be fooled into thinking that, that table salt glass video thing is true and then they'll self-rate that this actually did help me because now I'm not eating that glass. There's this challenge which comes down to back to a crisis of trust. Who do you trust to provide

these ratings signals, that distinguish between truth, meaning and fitness. It's not just about truth, it's about what provides meaning, what's also helping us survive, it's complicated. I mean, time well spent that phrase and that idea originally came from a new class of ratings that said not just what is it maximizing to spend our time, but what is of lasting benefit and value to our lives?

Tristan: Like a choice that on your death bed that you would say, gosh, I wouldn't take those hours back from YouTube for a minute. That totally changed my life. That was amazing. I laughed with those friends. I played the accordion, I learned how to play a few songs on the accordion from YouTube. I would embrace those hours with a big hug. There's so much good that can come here, but we need to have a totally different, almost like meta app that sits between us and all of this extractive garbage that helps us navigate just to these time well spent, lasting, humane, things that really recognize the things that make life worth living and also recognize we're human, we're naturally brilliant at certain things and making that happen more.

Aza: We often hear the line about platforms just being neutral parties. Of course is so intellectually dishonest, but I think most people don't realize we've baked in values to the very beginning. So Google search page rank was the sort of the algorithm let them rise, which is it determines how good a page is based on the number of not the content of the page, but based on how other pages view it to sort of social consensus of internet. And when they just let it run originally it just-

Tristan: Found porn.

Aza: It was finding porn, exactly.

Tristan: Porn is the number one thing that is most authoritative-

Aza: Right. Exactly.

Tristan: On the Internet according to that algorithm-

Aza: According to that algorithm so then-

Tristan: Unless.

Aza: Unless they're like, all right, we're going to have to seed it and started at MIT and Stanford. And so they were seeding the entire way that human beings, experienced our collective knowledge and they said yeah, these are values, there's some information which is better than other information and they baked it in. And now we just pretend like we're neutral because-

Guillaume: You can't be neutral. It doesn't exist.

Aza: Yeah. So where is your work blocked, right now? What do you need help with?

Tristan: And how can people help you?

Guillaume: Yes. So I think, the main blocking factor is this lack of public understanding of the problem. Thinking that Google has your best interest in mind. No. You should take a Google accountable. Like if people are ready to say no, like anti-vaccine promotion is not okay or Russian propaganda is not okay or not having transparency is not okay. So if this awareness that we need this transparency, we need the second accountability of algorithm is like the main roadblock. Once people understand that, then we can do all kinds of thing. We can give back more control to the user, either with Google doing it or startups doing it. I mean the startups can easily do it, but if there is no awareness, nobody's going to use their product. So-

Tristan: Yeah. So just to go in this plan about awareness, because I think a lot of people think, oh, raising that sounds important, but like it's not going to do anything that's not going to cause any change to happen. Let's talk about why and when this actually does make a change? So we just introduced this phrase of human downgrading, which is the connected system of problems. How it downgrades our attention spans, downgrades our civility, downgrades our common grounds, downgrades the quality of our beliefs and our thoughts, children's mental health. When we have a phrase that describes the problem, instead of talking about "Oh, there's some bad videos on YouTube", we're describing the problem not in a systemic way.

Tristan: It'd like be just talking, instead of talking about all climates changes, just talking about coral reefs all the time. People are like coral reefs are kind of important but is that such a big deal, versus if you can talk about climate change and how the whole system is moving together in this catastrophic direction. The first thing I think people can do is if you just have this conversation three times a day, human beings respond to public pressure. When there's three times in one day you hear it from a school teacher, you hear it in your design meetings, you hear it in your product meetings of people say, are we downgrading society? Are we downgrading the quality of people's beliefs?

Tristan: And not saying that in an accusing way, even though it sounds that way, what we're encouraging us to ask is just like we saw with time well spent, time well spent and the attention economy and technology hijacking our minds are three phrases that started to colonize the public discourse. And now so many people are talking in terms like that. It has led, along with political pressure, along with hearings, two huge changes. And in the past, YouTube actually has responded mostly when their advertisers get upset. So actually we might want to put out a call to the heads of Unilever, P&G on this podcast to be really aware of the systemic problem here.

Tristan: Right now, these guys, these CEOs of Unilever and P&G respond when there's a specific issue, like child pedophilia, right? Or there is an issue recently two weeks ago of YouTube recommending videos of how to commit suicide to teenagers. And when those issues come up, again because of researchers like you Guillaume, who raise it to the press, then the advertisers respond. But what we need is the advertisers to respond to the entire problem of human downgrading in a lasting and sustained campaign and effort. Because if they do that, then these companies can't continue to do what they're doing.

Tristan: And I think the whole purpose is we're in the middle of this kind of transition from the kind of fossil fuel, fossil attention age of the attention economy, where it's all extracted. We got drill on this race to the bottom of the brain, frack your mind, split your

attention to seven different multitasking streams because it's more profitable. That's the extractive era and that era is over. And we're moving now to a regenerative era of the attention economy, where we need every single one of these companies. Apple if you're listening, Google if you're listening, Android if you're listening, there's different players and different things they can do, to move away from human downgrading and move towards a more humane recognition of the vulnerabilities of human nature. And if we do that, we really do believe that that change can happen.

Guillaume: Definitely. There's so many, things that can be done. So we talked about optimizing different things, optimizing regenerative content, optimizing giving more control to the user. You could build in more metrics for the user to say, hey, this content was very helpful or this content hurt me in the long term. You could report much more kind of things on taking that into account. So there are a lot of solutions when people notice. So it's a bit like you're fighting the cigarette tobacco industry. It took so long like to raise awareness, but at some point when the media in the U.S. focused on raising awareness about tobacco people understood and then smoking became uncool. So common awareness saves lives and will save America I think.

Tristan: And just one last thing on that is just the urgency. So those issues, tobacco are huge issues and took 60 years to flip that around culturally. But in this case, when you realize the speed at which technology is evolving and that's super computer is playing chess, millions and billions moves ahead on the chess board every year. It's getting better. It's not moving at a slow timeline, it's moving at an exponential timeline which is why now is the moment, not later, now, to create that shared surround sound.

Aza: Even if I don't watch YouTube, I'm still surrounded in a media environment and people that do and-

Tristan: And everyone else thinks something-

Aza: Then.

Tristan: That's going to affect me.

Aza: Then it makes sense.

Tristan: Or they vote for someone else. Even if I don't, still going to vote for who I may vote for. We're all still affected by this. And that's, I think the main point to end all this, just that this is an issue we're all in this boat together, but if we turn our hand, put our hand on the steering wheel, we can turn it, just what we have to do.

Guillaume: Yes.

Tristan: Guillaume thank you so much for coming. This has been a fascinating conversation.

Guillaume: Thank you Tristan.

Aza: Yeah, it's been such a pleasure as always Guillaume.

Tristan: On our next episode Aza and I talk to Renee DiResta, a researcher who investigates the spread of disinformation and propaganda about how Russian campaigns have evolved with the Internet.

Renee DiResta: I hear a lot Oh, it's just some stupid memes. And it's interesting to me to hear that because they were running the same messages in the 1960s in the form of long form articles. So the propaganda just evolves to fit the distribution mechanism and the most logical information kind of reach of the day. And so in a way, they should be using memes in fact that is absolutely where they should be. And it's interesting to hear that spoken of so dismissively.

Tristan: A lot of us look at cartoons with silly messages in block letters on the Internet, and can't imagine that content like that would ever really affect anyone's opinion, much less their vote.

Aza: But Renee helps us understand that some of what looks childish and primitive on the Internet, is actually the result of sophisticated campaigns by foreign state actors.

Aza: Did this interview give you ideas? Do you want to chime in? After each episode of the podcast, we are holding real time virtual conversations with members of our community to react and share solutions. You can find a link and information about the next one on our website, humanetech.com/podcast.

Tristan: Your undivided attention is produced by the center for humane technology. Our Executive Producer is Dan Kennedy, our Associate Producer is Natalie Jones, original music by Ryan and Hayes Holiday. Henry Lerner helped with the fact checking. A special thanks to Abby Hall, Brooke Clinton, Randy Fernando, Coleen Haikes, David Jay, and the whole Center for Humane Technology team for making this podcast.

Aza: And a very special thanks to our generous lead supports at the Center for Humane Technology who make all of our work possible, including the Gerald Schwartz and Heather Reisman Foundation, the Omidyar Network, The Patrick J. McGovern Foundation, Craig Newmark Philanthropies, the Knight Foundation, Evolve Foundation and Ford Foundation among many others.