Jason Davis:

I'm a bit of a maybe technical nomad. I've been in the research and development space for about 25 years now. I'm a chemist by training. I spent about 10 years in pharma in industry doing research in drug

We're in this 24/7 news cycle where we know everything. We can look up everything on our smartphones, our mini computers in our pockets. I think there's this misconception, if you will, that this is the new age. And it is, I mean, there are new challenges, but we have always had some sort of misinformation out there when it comes to campaigns, whether you're going back to muckraking of journalism in the 1890s all the way through today. So give us a little bit of background, and Jenny or Jason, whoever wants to jump on this one first, the history of misinformation and campaigns and how we got to this point where we are today.

Jenny Stromer-Galley:

It's not the case that it's novel or new, that there are efforts at misinformation and disinformation happening around a presidential campaign. In the United States, presidential campaigns are quite consequential because of our two-party system and because the presidency is so impactful to the policy agenda in the United States and the power brokers that are around those political centers.

And so the difference I think right now is that we have a highly fragmented information environment. When you think back to say the 1950s, 1960s, we had a couple of television stations, there was a print newspaper that showed up in most people's households, they had radio. Journalists played an important role in helping to sift and sort what was good information from problematic information. And that vetting process has really, I think, gotten challenged in the current information environment that we have. We have multitudes of different places where people can get their information. Some of that is high quality information from people who are trying to vet and identify the good information from the bad information.

But depending on where people are getting their information, the quality and credibility of that information could be quite low. And so it leaves the public more vulnerable to say, state actors who are trying to engage in disinformation campaigns or US-based malignant actors who are trying to manipulate the public for their own ends. And so I think that's the challenge we face right now. It's not that it's a new problem, it's that the information environment has complicated the public's ability to sift and sort what is good quality information from effort that's meant to deceive.

John Boccacino:

And Jason, I feel this might be a good segue into your work, your involvement with the Semantic Forensics program. And this work is funded by the Defense Advanced Research Projects Agency or DARPA. Give us a little background into how your work with DARPA has helped further the research into that detection of disinformation and misinformation in the media.

Jason Davis:

Yeah, I mean the SemaFor program, as we like to call it for short form, is really about semantic forensics, which is the understanding of not just whether something is real or fake, but a little bit more on the why. What are the semantics behind it? What was the intent? Who was the target? That sort of level of characterization.

We got involved in this program. This is the fourth year that we've been working in this research area through the Department of Defense and DARPA. Because look, the US Department of Defense has been aware of this as a very significant homeland security threat and has been for 10 years now. This has been on the roadmap as a very serious challenge.

And look, I mean, it's not just a US-centric problem. This is a global challenge, and global actors are involved, global populations are the vulnerable targets. This is not just a US-centric problem. It is something that every nation state is struggling with, and in some cases adding to, in most cases adding to, right? Nobody's got clean hands here. This is one of the new, let's say, influence spheres that people are actively using as nation-states to further their own interests.

I would say what has changed is a couple of things. The social media platform scene has really changed the scale and scope of distribution, potential channels. It becomes very easy. The democratization of information creation and distribution has really made it so that anyone can create content and put it out there.

And there's been this weird phenomenon from a scientific perspective of a decoupling of credentials, that reputable vetting of information and speaking from a position of knowledge and expertise that Jenny just spoke eloquently to, to an influencer who has a large following and can speak on topics that perhaps they have absolutely no bona fide qualifications to speak to. And yet they have a powerful voice and influence on a large, millions many times, population that are followers and have created this illusion of digital trust and a relationship with them.

And so these have created an ecosystem that is far more vulnerable to mis- and disinformation. And so we as humans, our brains have not evolved as fast as the technology. And so we are still as vulnerable as we ever were to the same sorts of approaches at being deceived, intentionally or unintentionally. We have the same sort of vulnerabilities that we have always had, and the same tactics work on us pretty much every time. With this new digital landscape and digital speed and scale, we need some digital tools to help us protect ourselves from sometimes ourselves, sometimes that malicious information ecosystem.

So that is how we got to be where we are. And what we've been working on is developing those digital tools to try and identify both synthetic, or you can call it fake news or again, manipulated media. So humans can just as easily use old-fashioned techniques, old-fashioned like Photoshop and going in and just misrepresenting yourself with some very propaganda-driven narrative as a large language model or a synthetic generator.

So we have looked at two things really primarily from our perspective and activity in the SemaFor is how do we evaluate these detectors and really understand what they can do, what they can't do, and how well they do it? So that we can say with confidence, this detector works for detecting these kinds of fake, synthetic images at 98% accuracy. They don't do this and they don't do that, they're not a panacea, but here's what they can do. So use them properly and use them well. And that gives us some confidence in their results and lets us use the tools appropriately.

So that evaluation and then of course the development of the tools and the modeling of the threat landscape. So how do we create controlled versions of what we know is going on out there in the wild so that we can study them more deeply and really train and understand our capabilities on them. That's where our fit into the program lies.

John Boccacino:

It was interesting with social media and right away you knew that most people, I would hope, knew that it was a synthetic image, but when Taylor Swift came out, there was after the cat lady comments and there was the post of she supports Donald Trump. And everyone's like, "Oh, no, no. There is no way that a Swiftie would come out and support Donald Trump's presidency." But there were some people that fall for something like this.

And Jason, you mentioned this earlier, I want to pin this question towards Jenny about our brains being susceptible to these manipulations. Why is it that we can fall prey and be our own worst enemy when it comes to seeing through what's synthetic versus what's real? And why is it that we're always a step behind when it comes to these tricks that are meant there to deceive us?

Jenny Stromer-Galley:

Well, as Jason said, our brains have not been able to keep up with or adapt, if you will, to the rapidly changing information environment. We are cognitively hardwired to trust what our eyes are taking in and seeing. And it takes effort to engage in questioning whether what you're looking at is legitimate or not.

Danny Kahneman wrote a book called Thinking Fast and Slow. Kahneman and Tversky are cognitive psychologists who over a lifetime of research identified the ways that our brains process information. There are two systems to that. System one is the fast, quick heuristic thinking.

So if I'm scrolling through my social media newsfeed and a friend of mine shares a photo of Taylor Swift endorsing Donald Trump, then if I'm not thinking about it and I don't know that much about Taylor Swift, I'm not paying that much attention to the political campaign and my friend has sent this or shared this photo on their feed, I'm more likely to just accept it as true. Because I'm not engaging and thinking about, wait a minute, Donald Trump and Taylor Swift? Is that really likely that Swift would endorse Trump? So there's a set of heuristics that are at play.

System two thinking is that slower questioning, more actively deliberating. And social media, the way the platforms, the tech companies have designed Facebook, Instagram, TikTok, YouTube, et cetera, is to activate or work on system one thinking. So the buttons, the design, the notifications, the scrolling feed, all of those things activate or fail to activate system two thinking and instead are really more about system one thinking, which leaves us more susceptible to these manipulations.

And that does speak to issues around design. As a researcher who does pay a lot of attention to these design elements of digital platforms and utilities and tools to help reduce people's reliance on system one thinking, we know that the designs matter a lot. And so it requires, I think the social media platforms, it requires regulation to further provide cues and indicators that help people to know when something is synthetically generated versus not. And we really need better regulation in this space, but we don't currently have it.

So I said at the beginning, it's about our brains not being able to differentiate quickly when you're seeing something that's actually synthetic or manipulated information. But we do also need better institutional or structural responses to help us navigate this information environment.

Jason Davis:

And if I can, you bring up a fantastic point, Jenny. The social media ecosystem that basically 2005, 2006 came on the scene, this was like a mass scale experiment that just happened in the Wild West. And people forget what it was like before there was social media platforms and this information ecosystem. But all along, the last 15 years, these companies have been developing their processes specifically to engage us in the way that they want us to.

And there's a word that I want your listeners to understand, I thought it was pretty powerful when I heard it, and it's called friction. It's actually a term that they use to differentiate between that heuristic thinking and then that deeper thinking that Jenny was just describing so beautifully. The way it works is they want to reduce friction at all times because friction is that moment where you slow down, you take another thought, and then you make a decision versus that instantaneous, continuous feed of information cycling without that deeper thinking process.

So they have designed these platforms and maybe not maliciously at all, but rather monetarily. It's designed to do what they want it to do, and we are designed to eat it up. So it is unfortunately. But if your listeners are scrolling, that word of friction, how much friction are you currently engaging in? If your scroll is just going with no stopping, that means you are in a frictionless state and you are fully engaged in their platform.

So that's one thing you can think about is how do I slow things down for myself? Because the ecosystem is designed to make you have to make that decision. They're not going to make the decision for you. So that's just a terminology that I found pretty compelling, and it really makes me remember like, oh man, I'm in frictionless mode right now. Do I want to be? Let's make a conscious decision about whether I want that state to be occurring. Or if I want to stop it, I need to stop it because the platforms are not going to.

John Boccacino:

I love that term, friction. I love empowering people to think, don't just mindlessly be scrolling. Take the time to consume what you're seeing. And again, we're trying to equip our audience with some of the tools. So Jason, I'll start off with this question and Jenny, I want you to chime in after Jason has given his expert thoughts too. What are some of the tools that we can use and we should be using to determine the authenticity, whether it's an image, whether it's an audio clip, whether it's a video file? What can we do to better inform and make sure that we're not falling prey to these synthetic images?

Jason Davis:

Well, so the first thing to do is to stop and analyze. That's number one. And so once you do that, then anything, all these other tools that maybe I'll describe to you become possible. But if you don't stop, you cannot get to that higher order thinking on anything.

So I would say the first is particularly with synthetically generated content that's coming out, the weak links currently lay between modalities. So image generators make images really well. Synthetic large language models make text really well. They don't work together. So they oftentimes have a cognitive disconnect that for us as humans is very obvious. For AI, it's pretty challenging actually to connect those two worlds because they just live in different model spaces.

So the first thing I would do is if it's multimodal media, which most of the time it is, look at the image, look at the text and say, "Do these things really match? Are there inconsistencies between the modalities?" Let's, again, do some of that cognitive processing. If they don't match, then you absolutely should start to question. Maybe all the way back to two or three years ago when these synthetic generators were new, they had inconsistencies within a single modality. So let's say for images, you could look, oh, earrings weren't paired, or there was symmetry issues in the background. There were inconsistencies that you as a human could visually take in and make ... Something doesn't look right there to us.

As Jenny described, we're very visual. We've evolved that way. That's how we process threat information, and so we are really quite good at it. So AI has been training and training and training to give us what we want and what we So the first thido [lod(betwies. Sthin'e)2. a vi,tSo let's out, tan audng ant the immetryf tho t[modalitive)

come from in the outside world? And if you can quickly get to that, then you at least can make an educated decision about what level of credibility to provide with that.

So those are some basic tools that we as humans can still apply with our cognitive skill set because visually our optics are no longer the right sensors. We can't see an image and say it's synthetic or real. They're too good. The text is too well written. The large language models have gotten too good at giving us what we like to hear and how we like to hear it. We can't tell the difference anymore. In fact, we sometimes prefer synthetically written material to human written material at this stage in the game.

And again, audio really good, really strong, impossible to tell with our human ear. Video is coming, not here yet, but it's coming. So these are the tools we have left. We should leverage them with a healthy dose of slow level thinking.

Jenny Stromer-Galley:

I like slow level thinking. I like that a lot. The other thing I would add to Jason's really clear explanation of tells, if you will, in the digital space, the research is pretty clear that the best defense to misinformation is knowledge. So in other words, people who know a lot about American politics, they know the actors, they know events, they know a bit of history, they're current on current events are less likely to fall for these efforts at misinformation or disinformation than people who are.

The term that gets used in political science is low informed voter. So these are folks who, for a variety of reasons, are just not tuned in to the political landscape. And because they don't know about key actors, current events, where people align or stand on policy positions, they are unfortunately more likely to fall for these efforts at disinformation.

So the best salve in this wound is education. And of course, as an educator, that's heartening that we do

space is that strategic communication. So that's very intentional efforts to target pockets of the electorate. And we can actually get a little bit of insight to the targeting strategies, who is getting these ads, which we don't get when we see Facebook posts or tweets.

And so part of the interest this election season was being able to track not only what the candidates are doing on their social media strategy, but also outside organizations. And by outside organizations, especially in the Facebook space, that means everything from political action committees, the political parties, the typical actors that we normally would expect, but then also unclear, not sure who they are, what their motivations are that are running ads in this space.

So it's everything from individuals who are running ads to these secretive organizations that have no legitimate, or at least no clearly legitimate footprint in the political space. They don't have an address. They're not registered with the Federal Election Commission. We don't know who they are, but they're running ads in this space. And it's a stunning amount of money that's being spent. The Republicans are being drastically outspent by the Democrats.

So at the top of the ticket, the Republican ticket, so this would be Donald Trump and JD Vance, it's a 12 to 1 financial difference that they're being outspent, which is very surprising actually this election season because in prior election seasons, the Republicans were outspending Democrats, especially on Facebook and Instagram ads. And Facebook and Instagram matter because when you combine the two platforms, 60% roughly of Americans are still on Facebook, 40% of Americans are on Instagram, and this is adults by the way. And so it really touches a large swath of the public. So the ads matter. They matter a lot.

One of the things that we found much to our surprise, was a pretty large network of organizations, individuals, we don't know who's behind this, they're running scam ads targeted to people who are activated and excited about the presidential election that are basically capitalizing on their enthusiasm by turning over their credit cards and then they're getting scammed. It's a credit card scam.

That network, they've spent over $5 million in advertisements on Facebook and Instagram since September. Facebook is trying to take these organizations or take down these Facebook pages that are running these scam ads, but the scammers continue to stay a step ahead. So Facebook will shut down a page which halts their ability to run ads. Then they'll just create a new page and they'll start running ads on that new page.

And one of the cool things is we're able to track that those two pages which look completely independent, they actually are connected through some metadata that we're collecting through Facebook. So things like the postal address, telephone number, email address, the website that they link out to that starts the scam, they share those elements in common. So that's how we're able to track them. So the content can be different, their names are quite different, but we're able to track them through those shared contact elements, which is kind of cool.

John Boccacino:

How about taking the findings and being able to parlay that with the social media platforms to make them more transparent when it comes to, again, the sources of election advertising and messaging?

Jenny Stromer-Galley:

So there is a great deal of work that Congress is going to have to do to address this issue that we're facing. There is regulation, for example, around television advertising. So if you watch a TV ad, you'll hear the Tm. So aSo tthereo trjEMCdvertfeder1 Tyou cod 7(ca but the scthesll)Tsenator /asam badam thes /.4 (basicted to issue)T

I can imagine a set of potential policy changes that I think would help clean up this information environment. So better disclosure, setting some requirements in place that if anybody is going to run ads around political campaigns that there needs to be some sort of disclosure, there needs to be some indication of where they're coming from. One of the complications at the moment though is with the Supreme Court ruling that corporations are able to engage in speech around elections. It really complicates some of the legislation. But you can still have disclosure practices without forcing or changing the amount of money that's being spent or who is spending that money.

I do believe that disclosure matters. I think it helps people, going back to what I said earlier about design. So there are things you can do to help people sort out if it's legitimate or not, through disclosures and other kinds of legitimacy efforts. Having a website that is an active website that exists that people can go to and look at and see who's behind this. I think those are very basic things that we don't do in this democracy that other democracies do that I think would actually help clean up some of the nefarious actors that currently exist in the political space.

John Boccacino:

I know that we've equipped our audience hopefully with some really pertinent skills and tools to prepare themselves for handling misinformation, disinformation. I do want to give Jason a question for you that I know you'll have a good answer for our audience. We've talked about skepticism, and it can be easy to leave this conversation thinking, "I can't trust anything that I'm seeing." Why is it important that we as a society don't fall into that habit of thinking, "There's nothing I can trust. It's on myself. Everything's out there to deceive me"?

Jason Davis:

The answer, if you go that route, is to become completely oblivious and uneducated about the world around you. And that's a dangerous position in and of itself. So we really cannot just shut ourselves off from information and think that that is a viable solution and a way to live. We are going to be exposed to it, but what we do have, and we should educate ourselves, and some of the tools that Jenny mentioned I think are really important, diversity of news source.

That's how you can engineer in protections for yourself. Intentionally seeking out polar news sources, even though there's one side you probably don't agree with, you should at least understand what their