ICANN82 | CF – NARALO Roundtable on AI and DNS Abuse
Monday, March 10, 2025 – 15:00 to 16:00 PST

MICHELLE DESMYTER: Hello and welcome, everyone, to the NARALO Roundtable DNS Abuse and AI Combatting and Enabling Fronts. My name is Michelle DeSmyter and I am the Remote Participation Manager for this session today. Please note that this session is being recorded and is governed by the ICANN expected standards of behavior, and ICANN community anti-harassment policy. During this session, questions or comments will only be read aloud if submitted within the Q&A pod. Interpretation for this session today will include English, French, and Spanish. If you would like to speak during the session, please raise your hand in Zoom; when called upon, virtual participants will be given to unmute in Zoom. Onsite participants will use a physical microphone to speak. Please state your name for the record, the language you will speak — if speaking a language other than English — and please speak at a reasonable pace. And with this, I will hand the floor over to Greg Shatan, NARALO Chair.

GREG SHATAN: Thank you everyone, and thank you for coming to our roundtable on DNS Abuse and Artificial Intelligence. A very timely topic — a combination of two very timely topics, of course. And without further ado, since we have a panel of luminaries, let me introduce

them. First, directly to my right, Jeff Bedser is a seasoned entrepreneur and leader in cybersecurity, investigation, threat intelligence, and DNS abuse mitigation. He is the CEO of CleanDNS, a company that is on a mission to clean up the internet for good — for better or worse — by detecting and stopping malicious DNS activities. In his pas life, he was a private detective, and he uses that expertise to protect internet end-users, the people that we care about here in At-Large. He's on various boards and committees on internet governance, stability, and jurisdiction, and has 25 years of experience in the field, founded iThreat, a threat intelligence company. And he has been serving on the SSAC at ICANN since 2007, and was formerly on the PIR board. So, Jeff, thank you very much for being with us. Tim Maurer is to his right. Tim is currently the Senior Director of Global Cybersecurity Policy at Microsoft, where he collaborates closely with Microsoft's internal teams and stakeholders, and external stakeholders. Before Microsoft, Tim was with the United States Government, first as a Senior Counsel for Cybersecurity, Emerging Technologies for Homeland Security, and subsequently, on the White House National Security Council. I believe that was under President Biden. And has spent a decade working on cybersecurity and tech policy, and is the author of *Cyber Mercenaries: The State, Hackers, and Power*. To his right, we have Ana Neves. Many of you probably know Ana as the Portugal GAC representative. Ana wears a number of hats: She's a member of the current IGF MAG, which is a board of heavy-hitters that keeps the IGF running. She is the chair of DNS.PT's advisory board, and leads the Internet Governance Office at FCT, FCCN MECI in

Portugal. She's also the Vice-Chair of the United Nations Commission on Science, Technology, and Development. She has over 30 years' experience in digital science and of technology development in public policy, and was recognized in 2018 as Portuguese CIO Public Sector of the Year and European Digital Leader. So, we have a panel that makes me feel obscure and underachieving. So, without further ado, I would like to kick off the discussion. So, if we can go here to the first slide — before we go any further with this, I'd like to mention that this entire presentation was written by Microsoft Copilot. I'm not joking. There it is. So, instead of "God, be my copilot," Microsoft is my Copilot. I hope you appreciate that, Tim.

TIM MAURER I'm ready, I'm ready; the human backup is here.

GREG SHATAN There you go. So, according to Copilot, DNS abuse refers to the misuse of the domain name system for malicious purposes, such as phishing, malware distribution, and other cyber-threats. The kind of official ICANN definition of DNS abuse has five categories, maybe one for each ICANN geographic region: Malware, botnets, use of botnets, phishing and pharming, and spam — but only when spam serves as a delivery mechanism for one of the other four. So, spam per se is not considered to be DNS abuse. Those of you who have been around ICANN for a while can remember, we had several years of… discussion is a good term for it, about the limits and definition of DNS abuse, and rather than discussing it any more, or at least

living with this definition of DNS abuse as we go forward, since this certainly gives us enough problems. As you know and I know and Microsoft Copilot knows, AI and Machine Learning are increasingly being used to combat DNS abuse by detecting and mitigating these effects more effectively. Next slide, please. So, I'll let you read all this stuff on your own time; we don't need to hear it. I'd rather hear what our panelists have to say than what Microsoft Copilot has to say. And the design, of course, was chosen for me by Microsoft PowerPoint, and the picture on the front was also chosen by Copilot. I'm actually a very creative person but now I've decided just to let Copilot do all that stuff. In any case, I'd like to turn first to Tim. I'm just going to say — for one more minute, of course, DNS abuse is also being manifested by Artificial Intelligence, and there are several examples in this PowerPoint as well of the ways in which AI is creating new forms and mutating older forms of DNS abuse. So, we have kind of a battle of the Artificial Intelligences, for good and for evil, and our three panelists will discuss both the general cybersecurity and AI interface, how that affects DNS, and anything else that's on their mind; so, kind of give a broad overview. Tim, who is the only non-ICANN insider on our panel, will get to kind of kick us off.

TIM MAURER Thank you very much, Greg. And it's a pleasure to be with you all here. I was just talking to Ana, I think the last time that we had a chance to meet was prior to the pandemic, which is I think also the last time I had a chance to attend an ICANN meeting. And it's a particular pleasure to be here with all of you today, because this is

a community that we at Microsoft are very excited about, to hopefully help us use AI and maximize its potential for the defense of the internet, and make sure that it is used to actually help achieve what we hope the technology with unlock, which is further human progress, and also make sure that we use it in ways that it will help us advance existing objectives and goals that we share. To Greg's point, I just joined Microsoft a little over a year ago, so it's been a fascinating experience for me to now be at the company and at the forefront of AI innovation. And the way that we are thinking about AI more broadly, but then also specific as it relates to DNS abuse, is that we are cognizant of the fact that AI has been around for a while, depending on how you define it, and that Machine Learning and predictive AI, and analytical AI, has been used for a number of techniques and ways to already mitigate DNS abuse. So, as Greg alluded to with Copilot, generative AI is hopefully going to help us unlock further potential through a number of ways, and actually make also the way that we can help mitigate DNS abuse more accessible to a broader set of defenders by breaking down barriers that may have existed before about skills level, and using generative AI, in particular Copilot, and other tools that we have to make it easier for the defender to use tools to mitigate abuse. One thing I will just upfront and then I'm going to hopefully make this as interactive as possible, with the panelists and all of you in the room, is that the company is looking at AI in the context of security and the context of DNS abuse through three lenses. The first one is, how do we maximize the use of AI for defense, and that is how do we mobilize customers in communities like yours, in the ICANN

community, to adopt the technology quickly; and the reason I highlight speed is, I can guarantee you, those with nefarious purposes will be among the fastest to use the new tools to figure out how they can use it to advance their existing malicious activity. So, the faster all of us can try to advance AI and use it for defense and mitigation techniques, the faster we can make sure that we can catch up — not just catch up, but also outrun the threat actors that are hitting organizations around the world every single day. The second piece I would highlight is that the company is early on very focused on figuring out what steps we need to take such that threat actors cannot take advantage of our AI products and services. So, a year ago, the company released a policy that made clear, if we detect that a threat actor the company is tracking, of which you can imagine there are many over the years that have in the past conducted an unauthorized access or otherwise came to our attention, and we noticed that they are using AI products or services, we will take action to terminate their account. It doesn't matter if they are using AI in a productive sense, we don't want criminals to become more productive, so we will take action against that and that is a policy that came from the leadership, not in response to regulation or government request, but was a reflection of the leadership's commitment to make sure that we empower the defender and try to make it as difficult as possible for the threat actors involved. And then the last piece, in terms of just overall framing, is that the company is focused on what is it that we need to do on our end to make sure that the technology isn't used, as we think about Azure and our cloud services, and how can we

also protect our AI products and services, which, as you can imagine, are being targeted by other threat actors who would like to know the insight of how we use them. So, let me just use that kind of as a general framing of three dimensions; I know there's a lot to discuss when we go into the details of DNS abuse, but with that, over to you.

GREG SHATAN Thank you very much, Tim. And I noticed of course you mentioned DNS abuse defense a number of times — so, we're lucky to have with us Jeff Bedser, whose life is entirely defined by defense and cleaning up the DNS. I'm not going to call you the new sheriff in town, because that's probably somebody else's trademark at this point, but in any case, Jeff, I think-

JEFF BEDSER I do have a badge.

GREG SHATAN Oh, well there you go. You have a badge. We don't need no stinking badges. So, in any case, Jeff, I'd like you to kind of dig a little deeper. And I know you have a slide deck, and probably is not written by Microsoft Copilot, might have actually been written by you. So, without further ado, I'll let the slide deck and you speak for yourself.

JEFF BEDSER

Thanks, Greg. And no, this one hasn't been but I cannot deny that I have used AI successfully to do many decks. So, I'm going to capitalize on the fact that I spoke to this group, feels like a week ago already, yesterday? The day before, Sunday? About DNS abuse in a general sense, and I think that this presentation I've put together is about how CleanDNS is using AI processes to improve in that fight, just as Tim said. And I mentioned I think in my presentation the other day, was that if you're thinking about it now, they already thought about it, did it, and probably moved on to something faster, because they're that far ahead of us. And if we're not going to be using the emerging technologies at the same rate and pace for defense, we're going to get so far behind so fast it's going to leave everyone's heads spinning. So, what I'm going to be talking about is how we've been actively deploying API models into what we do, because again, it comes down to, like I said the other day, there's a significant volume of abuse that we don't know about because it's not being reported. What we're seeing in all these active reports that are being done about five to seven sources that most people run the reports on, but that's the tip of the iceberg. And literally, if you don't see it, it's not reported, and the majority of people don't report the phish or the smish or whatever else they're getting; so, if we don't know about it, we can't act upon it. So, the best solution is to detect it quickly, and when you detect it quickly, you can mitigate it quickly. And that's the space that CleanDNS sits in — we sit between the companies that detect the abuses and then the industry, our industry, between the registries, registrars, and hosting companies who need to mitigate it. And our

goal is constantly to speed up not just the reporting but the detection, so we can get to the point where the process between detection and mitigation is seconds, whereas I don't know if anybody has read that tome that I was the Work Party Chair for SAC115 on interoperability and DNS abuse handling or management, but we found in that study several years ago that the average time was 96 hours between report and mitigation; and I think we can all guess how many people click on a phishing link and 96 hours, and how many victims that accounts for — and how many fewer, if we can get it down a couple of seconds versus 96 hours. So, let me get into the presentation. Next slide, please. So, SAC115 was the template from which I decided to build CleanDNS, because many people in the industry pointed out to me that while it was a really good academic paper, unless somebody built the solutions that were talked about in the paper, it wasn't going to exist. And CleanDNS was built to match that. I'd like to call it a virtue cycle, and I'll give credit where credit is due, Byron Holland from CIRA used that term with me when we were talking about it — we're trying to not just solve a problem, we're trying to make the reporters better and we're trying to make the industry better by solving the problem in the middle of, you found it, now what kind of evidence is required to act upon it, and then shorten the cycle of killing it. So, it's an aggregation model to basically escalate that process. So, let me get into what we're doing with AI, and that's the next slide, please. So, the first one is easy, and it's one of those places where you can quickly see that the AI is better than a human. So, everyone has been phished in this room; anyone that hasn't,

please explain why you're at a technology conference and you don't have a phone. No? OK. Phishing is based on lures, and the lure is, some company, to make you believe that you're going to something it is not, whether it's a bank, an ecommerce platform, whether it's your telecom, whether it's your local doctor, there's a lure involved. And that lure is almost always an image that makes you believe it's that bank, it's that ecommerce platform, etcetera. Well, for humans doing analysis of phishing, they don't know every bank on the planet, they don't know every company on the planet, they don't have knowledge — they may have been doing it for 30 years and they're never going to know it all. The AI, [IAI] systems can pull from databases and systems that have those image files for quick comparisons. So, Image Analysis can give you the ability to validate the logo, the Image Analysis can look and see, is there an attempt to grab credentials, is there an attempt to grab personally identifiable information? And when I say Image Analysis, it's basically a screen capture, so the landing-page of a phish or a pharm or a fake shop or a smish. I'll stop going through all the acronyms. But you're basically able to do that screen capture and you can analyze it for the content, and then the AI is very good at saying "yes, this is structured in a way that is likely a phish." Sentiment Analysis is based on looking at the text and looking at other components of it to say — I made that text too small on the screen, I apologize; we can share the presentation afterwards if anybody wants it. Maybe I can read it on my screen for you, hold on. It says "Analysis of inbound emailed reports related to content and intent of the reporter." So, the industry for the last 20 years has

basically been an email-based reporting mechanism. And of course, the problem with that is it's unstructured data. So, Sentiment Analysis allows you to take that unstructured email data and try and figure out what's happening. Humans are, by default, storytellers and when you had something happen to you, you want to tell the story about "how I got catfished" and "why I thought this person thought I was so attractive and that they wanted me to send $50 so they could come visit me," and the stories are fascinating to read but take a lot of time. And what you're really looking for in that text is what is the URL or domain involved, and what are you accusing it of. Ah, you were phished and that's the URL — that's all I care about. The rest of the story is irrelevant to a mitigation effort. So, the Sentiment Analysis helps us get there. And in a contextual relationship — again, putting on my readers, I apologize — is the vector embedding the similarities between the blocks of text. So, we're looking for patterns where, for example, spam or phishing, is this the same campaign from the same cybercriminal, because they're using the same structure, format, etcetera, so we can develop an understanding of larger groups. We're not taking down one at a time, can we use the contextual relationships to figure out "oh, there's 10,000 of these, let's take them all at once because they're all related to the same cybercrime based on the timestamp" the contextual relationships between the data. Next slide, please. So, I got ahead of myself but logo entity detection is part of the Image Analysis. PII detection is, again — one of the other keys is that a lot of domains that are going to be used for abuse, or have been used for abuse, they put it to a parked page before they use it

for abuse, and sometimes, they'll use it for abuse, switch it back to a parked page, so when the good guys are looking at it, they're like "oh, it's just a parked page, there's nothing there," and then they wait for that traffic to go in and turn it back on again, and now it's on a phishing page again. So, detection of whether it's a parked page versus a phish is an important component of using Image Analysis, because you don't need a human to do that; again, we're just shortening the span and speeding up the timing. Next slide, please. So, potential pursuits we're looking into is the finetuning of this whole process. So, it's great for phishing, but what other harms can we do here? Because again, DNS abuse as defined under the contracted parties and their obligations are the ones that Greg spoke to, but, everyone in this room knows there's plenty of other online harms, and those plenty of online harms need to be mitigated. And most of the parties that represent those in the contracted parties' house and the ccNSO act upon many other types of harms that are not required to act upon — they do it because it's the right thing. So, how do you finetune an evidencing package that says "this type of harm, when presented this way, is a harm," so that it can be mitigated? Malicious code detection, that's difficult because if you have to process the code to determine if it's malicious, you're putting your systems at risk of being compromised. So, using the AI to take and look at the patterns to say "which malware family is this? How is this targeted?" is another place where AI can be used to fight that. And then lastly, the generative AI component of this — we've been using it to say "structure the outbound communications to the parties affected";

if it's a registry reporting to a registrar, we use it to generate the text to explain what we've seen, what happened. But we use it to create templates, not to create the actual communications, because anybody that's played with AI, as Greg pointed out and [had to admit] the story to him, you've got to be very careful because unrestricted generative AI starts to fantasize. And I actually had it write a paper for me on all the entities working on DNS abuse, and it came up with some great entities that I researched and I researched the footnotes, and none of those entities existed. The websites they referenced from didn't exist. So, you've got to be very careful in generative AI to control the environment from which restrictions you give it. Next slide, please. So, pitfalls — transparency. There's nobody that works with me that doesn't understand we're using these technologies, and we don't want to pretend we're not. You have to watch your data biases, and that is also a legitimate concern when it comes to which datasets you apply. If I start bringing in datasets from reporters that have a lot of false positives in them but don't know they're false positives, I can corrupt my whole dataset on validation. For example, if there's a whole bunch of image files that actually aren't ecommerce or banks, and they get into that dataset and now they're being confirmed as such, that bias can be problematic. False positives and negatives still require — at what point do you kick it out for human review? And apparently, that's Tim. So, I'll need your email address for that. But of course, sensitive case management about the data and the controls of those data, typically for privacy

purposes, etcetera. So, I think that's the last slide; so, I will turn it back over to you, Greg.

GREG SHATAN Thank you very much, Jeff. Very interesting and timely. Next, we'll turn to Ana Neves. Ana, you can say whatever you want.

ANA NEVES OK, thank you very much. So, whatever I want? Well, it's very good to be here in Seattle, and in particular here in At-Large; it's always a great honor to be here At-Large, and because since the beginning of my life in the GAC, I always defended that GAC and At-Large should really work together, because we should enhance our efforts in the public sector and in the end-user sites. So, what can I say after the Microsoft and Jeffrey said and explained DNS abuse? What I can say about Microsoft is that it's very good to know that you have a policy to mitigate DNS abuse, but what about other companies that don't have that policy, and that policy came from who? And that policy is envisaged by whom? And can we rely on companies to have such policy, and can we rely — why? So, from the public sector perspective, that's what I think that I was invited for, to be part of this Roundtable, is to strengthen the collaboration between public and private sectors, and to have more partnerships, and to see with the regulatory bodies what can be done, and to have that always in mind, that cybersecurity, to have that in mind in governance, when they are regulating or thinking about public policy, to have always DNS abuse in mind. And so here, multistakeholder efforts are really important to ensure

responsible AI usage in DNS security. Another important point that I would like to highlight here is the development of AI governance frameworks. So, I think it's very good to discuss these here at ICANN, and together with At-Large, which could explore policies that ensure AI driven security aligns with global internet governance principles; and this is a very good year to do that as we are reflecting on the WSIS+20 review, how to streamline this process with the Global Digital Compact, and what should we really envisage in all these processes and negotiations. And the last thing that I would like to highlight in this, my first comment, is to empower end-users. So, AI-driven protections should be designed with transparency and allow individuals to make informed choices about security and privacy settings. So, it's always about to have good sense but to have end-users well informed, so, capacity-building is always important, privacy is always important to underline in all our efforts. So, I would say that strengthening collaboration, having more public-private partnerships, developing AI governance frameworks and empowering end users is my first three items for this discussion and for this Roundtable. Thank you very much. Greg.

GREG SHATAN

Thank you, Ana. Can we go back in the slide deck to where I stopped before, which was basically the first real slide here? So, let's go to the next slide, please, just to cover some of the things that our good friends at Copilot have found, just to put a few more facts on the ground. So, we're seeing AI-powered phishing attacks, which are using AI to personalize phishing, almost making it a bit more spear-

phish like, mimicking trusted contacts, bypassing security filters, improving the syntax and grammar, unless that's being used to — unless the mistakes are there to make sure that people who are smart don't read them. And as well, AI can be used to prejudge credentials, test thousands of stolen credentials in seconds, enabling brute-force attacks and basically the repetitive nature of these attacks, this is a volume business; and so, AI can enhance the power of volume in bad actors' threats. Next, malware can also be enhanced by AI, with code that develops dynamically to evade detection, and helping it to hide from antivirus software, which, of course, is looking for patterns that it knows and learns, also of course using AI. So, again, kind of the battle of the bots. Next slide, please. We also can see reconnaissance using network scanning tools using AI at speed and at scale, identifying weak points in security infrastructure, kind of automated penetration attacks as well. And then, social engineering again can be made more dynamic, more real, yet fake, to bypass traditional filters, again, using AI to tweak things so they appear less like obvious threats. So, this clearly indicates why we have to be using sophisticated and constantly improving defense features to keep up with all of this. Next slide, please. So, some of those solutions, of course, are AI-powered DNS security tools using zero-trust architecture with a stricter verification. I think all of us have noticed more use of multifactor authentication, more use of different types of authentication and layers, even just in the last six months to a year, just to try to keep up with the threat levels. Next slide, please. So, threat intelligence platforms need to keep enhancing; Tim has

already — or rather, Jeff has already alluded to that. Of course, there's always the human factor, we still need to train real people, employees and end-users, general public, on how to recognize phishing and social engineering attacks. Awareness needs to be kept up constantly — I just did my firm's annual cybersecurity training, and it's probably already out of date. Next slide, as well. As a lawyer, I can appreciate the need to continue to enhance legal frameworks; the law never moves as quickly as AI, of course, but trying to keep both legal and ethical rules in ways that govern AI more responsibly. And one of the things I'm curious about is whether the companies that are working on AI are aware enough of the DNS abuse universe, as opposed to all of the other universes that are using AI. Our space, while it's important, is not necessarily top of mind. So, whether AI companies are doing enough to work with the rest of the ecosystem. Last but not least, of course, continuing more proactive security measures, heightened paranoia, is also critical. Next slide, please. This is Jeff's bio, at least the one that was given to me by Microsoft Copilot; you can read it at your leisure. It seems pretty accurate, didn't make too much stuff up. Next slide, please. I first asked for Tim's-

TIM MAURER There was one error you should point out.

GREG SHATAN Oh, no? What was that?

TIM MAURER Go back a slide.

GREG SHATAN Go back a slide.

TIM MAURER Look at the year on the bottom second paragraph.

GREG SHATAN Oh, go back — oh, the year 2071. Well, actually, that's because I — there was a footnote there for one which I didn't remove. So, that's a human mistake. Next slide, please. This was the first attempt using AI to get Tim's bio; it did not mention Microsoft at all. Again, hopefully more or less accurate, but again, not up to date. Next slide, I put Microsoft into the prompt. Next slide. I added Microsoft to the prompt and voila, we find out he's the Senior Director of Global Cybersecurity Policy at Microsoft, but I wouldn't have known that if I hadn't put it in the prompt. So, Microsoft Copilot doesn't know Tim well enough yet. But I can see him taking notes, so that'll probably be fixed. Next slide, please. I wanted to know more about Tim and his hobbies, so hiking, cycling, photography, reading. He has diverse interests and commitments to personal growth and wellbeing. At least, that's what Microsoft Copilot thinks. Next slide, please. This is Ana's biography; again, doesn't really hold a candle to the bio that she gave us, but it doesn't seem to be completely off-topic. But next slide, my first attempt resulted in a completely different Ana Neves, Ana Luisa Neves. I didn't use her middle name in this, but this person also is very much in our

world in a sense — she's the Director of Global Digital Health at Imperial College in London. And I hope the two of you met at some point, and if not, maybe you should look each other up and you probably have a lot of interesting things to discuss about our digital future. Next slide, please. This is me, if you want to know anything about me. It's a little old, it says I attended over 10 ICANN meetings — the real answer is over 35 ICANN meetings. And I'm no longer the President of ISOC New York, I've demoted myself to secretary, still on the board. Everything else is more or less true. The next slide, please. That's it — this is another picture of Seattle in the fog, as chosen by Microsoft Copilot. So, let's get back away from this slide — you can leave this one up, it's kind of pretty although somewhat apocalyptic looking. Let's get into the discussion of the offense and defense of AI. And I'm wondering, we'll start with you Jeff, what are you anticipating in the next stages, and when you look at how AI is changing, what is that going to do to the DNS abuse area that we're in?

JEFF BEDSER That's a great question, Greg. And man, I hope no one is listening to this and coming up with ideas out of me. So, I think the space we're going to see the most of it is in identity management and false identities. The ability for AI to create almost perfect national, local identity cards with the identities of real people on it, but maybe an alternate photo and such, so that it makes it almost impossible to validate a real person from a false person, is a big concern. And when you think about the fact that you can now say, "OK, AI" — not his AI, because it's stopping this stuff, I'm not being

sarcastic, it does stop this stuff — "AI, I want a credit card number that has not been used yet, that's been stolen from a breach. I want to match it to an ID for that same person, with their actual home address on it; and if they're domiciled in the state of Pennsylvania in the U.S., I want a driver's license that matches the Pennsylvania driver's license model, and I want to put my photo on it in a fashion that looks like a driver's license photo. So, I guess no smiling. And then generate that image, so that when I want to go buy a domain, buy some services with that stolen credit card, it passes all the checks for Know Your Customer." The credit card has not been flagged before, there's a perfect ID to match it, and that was all generated in, I don't know, what do you think, four, five seconds. That, when it comes to cybercrime, cybercrime has to be fueled by something; it's mostly fueled by stolen credit cards and comprised bank accounts, and you get to that money with basically modified stolen credentials. And the way to get those is basically those hard credentials that validate who a real person is. So, I think that is one of the biggest ones that we're going to see emerging, that's going to really kind of change the landscape quickly.

GREG SHATAN

Thank you, Jeff. Tim, I'll ask you kind of the same question, but also ask you to kind of think about the earlier question I asked, which is what are the developers of AI doing in particular, are they aware of the DNS abuse space versus all the other potential things that AI could be used be for, good and bad?

TIM MAURER

The answer is yes, partly because it's in Microsoft's self-interest, given that we also run a significant cloud infrastructure and rely on making sure that overall, the internet works in a way that it's not abused by threat actors. To just build on the example that we heard about, the way that the attackers are using it, I think there's some good news in it, in that since the advent of generative AI and the various models that have been put out, there were some that predicted a much more apocalyptic kind of scenario and what could unfold. We have not seen the resurgence of worms that we saw in the early 2000s that you could have imagined by generative AI writing new code that would have prevented avoiding existing detection mechanisms or just the filters. What we've mostly seen is existing threat actors incrementally adopt AI for known TTPs and advance their objectives. So, that's not great, obviously, but in terms of the spectrum of possibilities, I think there's some good news here in terms of what we've seen. And then, I do think there is some interesting lessons learned already, looking back at the past couple of months, in terms of how defenders are using it. And as you demonstrated with your prompts, it can be fun in terms of seeing what it produces, but it also highlights the importance of how you actually use the technology, what prompts you are using, how specific are the prompts; and to translate it now into defensive techniques, if you for example use a Copilot for security or any of the AI products and services, to produce specific reports when an anomaly is detected, and to just use it to automate more for example flagging that a specific anomaly has been detected, to produce an executive summary. We also want to make sure that we

make sure to instruct the model to err on the side of including more rather than less potential anomalies so that we capture the false positives, because what humans should be doing, then, is reviewing what gets flagged, and we prefer capturing more false positives that we can then discard, rather than the model potentially reducing the number of flags that may include more false negatives than anything else. So, I do think one, the human does remain important; the way that we can leverage AI is ideally use it in a way that it helps us use the humans more effectively. And what, to your point about scale, makes me somewhat optimistic is, despite the concerns that you flagged about, and particularly that identity piece which I completely agree with, is that one of the main challenges we've had on the defensive side has been a lack of skills and a lack of enough people trained to do certain tasks. Which, if we can get AI to use at scale to reduce those barriers of entries, and use it to produce reports that are more easily consumable, even to folks who are less skilled in cybersecurity, then we may have an outside effect in terms of the overall net assessment of whether AI is helping advance defensive work or offense. Just one quick note to Ana's point, that I didn't want to leave uncommented, about the importance of engaging between the private sector and the public sector, I think looking back at the last few years, it's been fascinating to see how post-release of ChatGPT and the New York Times kind of front page, how quickly regulators, be it with the EUAI Act, or in the U.S. government, the case of the various policies that were in place, you had a very rapid implementation of new regulation and laws and policies. Frankly, to the extent the

company has had to make sure that they were catching up. And with the emergence of AI Safety Institutes, which again, I was supposed to help make sure that we have mitigations in place that you can't just type in "oh, tell me this credit card number," I've actually been quite impressed with how quickly governments and regulators have put frameworks in place within a very short amount of time, and how they've engaged with the private sector, be it Microsoft, be it many of the other companies, to try to put some frameworks in place that are risk-based and focused on what we should be paying attention to.

GREG SHATAN Thanks. Ana, I'd like to get your perspective. And particularly, I'll ask for a European perspective, because the three of us, among other things that we are all in common, we're all very-

ANA NEVES Yes, very American.

GREG SHATAN -we're sitting in America all the time. Which, of course, is changing massively, as well. But I'd be interested to put some different perspectives into the mix here.

ANA NEVES Yes, absolutely. So, in Europe, I think that we understand that the threat landscape for DNS abuse is rapidly evolving with AI and Machine Learning. So, we are trying to develop guidelines for the

ethical use of AI in DNS abuse prevention, while addressing privacy concerns related to AI-driven DNS traffic analyses. So, we have this report done in the European Union, the European Union's comprehensive study on DNS abuse, which provides recommendations for registries and registrars. It is from, I think, December of 2022. And it is interesting to see that domains like .DE and .EU are the domains that are less… less abusive. So, where DNS abuse is not so strong. So, this is evidence that some work and the technical implementation and this kind of work together between the public and private sectors is very important. It's very important as well to work together with researchers on AI, because those are the ones that can help us in preventing. And so, to have this component of prevention ahead of what can be happening, is very important. So, again, multistakeholder, it's so important in all these digital processes and DNS abuse is another one. So, the importance to include academia researchers and developers of AI in all these processes are tremendously important. And so, only which this kind of involvement and to listen all these people and entities and different stakeholders involved, we can achieve different layers of policies, being policies for companies, being public policies to address DNS abuse. That's my take for now. Thank you.

GREG SHATAN Thank you, Ana. We have a lot of people in the room, so I think it's — and virtually, as well — so I'd like to turn to Q&A. So, if you are remote, please put your questions in the Q&A pod, and they'll be read aloud by Michelle DeSmyter. And if you're in the room and

you're on Zoom, please put your hand up. And I see Joanna's hand is up. And if you are not on Zoom, just raise your hand. And if you're not at the table, we have a roving mic which will come to you, so don't be shy if you're not sitting near a microphone, the microphone will come to you. So, we'll start with Joanna Kulesza.

JOANNA KULESZA

Thank you, thank you very much, Greg. This is Cindy, Joanna Kulesza for transcription purposes. Thank you for all the presentations. It's wonderful to hear about the link between AI and DNS. I have a question that piggyback rides on Ana's intervention and that last comment. I think it might make sense for me to target this at Tim — thank you very much for that very comprehensive review, you've sort of addressed this. I'm not sure if you're comfortable answering this with your Microsoft hat on, so feel free to choose how to address it. We do have a shared research background, so if you want to speak about global cybersecurity policies, that's perfectly fine as well. I'm going to start with a little background for that question, but then I would love your feedback. On one hand, Microsoft was vital in supporting Ukraine after the unprecedented Russian aggression — Microsoft intelligence, cybersecurity, threat analysis was vital for the Ukrainians to prepare their defenses. The situation might change shortly; it seems that the information sharing was stalled, now it might be coming back. I think this is a very clear indication of how national regulation could target attempts to ensure cybersecurity, be it at the DNS abuse level or be it general with regards to sharing information. On the other hand, Ana mentioned, and you indicated

this as well, that there is enhanced regulation on AI, the AI Act included. Senator Cruz, in December of last year, said that the AI Act stalls innovation and could be considered foreign interference into AI development in the U.S. Now, this brings me to my last point, and the question itself. Our good friend, Marietje Schaake, published a book, the IT Coup — I'm certain the Tech Coup that you might be aware of the concept there — and recently speaking about it, she mentioned that this could be the quantum moment for researchers who do not feel very comfortable researching in the U.S. to move to the EU. The EU, she said, should roll out the red carpet, sort of in a reverse scenario that we witnessed in the mid-1930s. So, with your researcher's hat on, I'm not sure it's a comfortable question for Microsoft to answer, and I wouldn't pose it in this setting — what do you think is the role of regulation in developing AI for cybersecurity? And I will turn this into a policy-related question before Jonathan puts me in my place — what can the end-user community do, maybe working together with a governmental advisory committee, to make sure that we use AI to protect the individuals from cybercriminals? I believe it's a diplomatic way to put a question, and I know you can clearly identify the geopolitical challenge I'm trying to address. Whatever thoughts you might be at liberty to share will be appreciated. Thank you so much.

TIM MAURER Thank you, Joanna. And also, good to see you again, which I think also goes to pre-pandemic days and before my time in government. I'm happy to answer this in a Microsoft capacity, because the

company has been very clear with respect to some of the geopolitical challenges that we've encountered, and is very committed to cybersecurity up front for the company, for the customers. There's the Secure Future Initiative, which has put security back front and center for the company all up. Let me address the last point first about what this community can do. I think we are at a really pivotal moment of technological innovation where the process we are making with Artificial Intelligence, and Microsoft has been rolling out new products and services on a fairly regular basis, and plans to continue to do that, is for this community to help us figure out how best to use the technology for good, and for, in this context particularly, their defense. And this really depends on, you can bet that the people who are focused on monetizing nefarious actions will be quick to test out how these new tools can help advance their objective, so we need to make sure we are fast in adopting the technology. And if you identify areas where you see vulnerabilities, where you consider there to be research questions that a company like Microsoft should be focusing on, then please do share those with us, because we are in the same boat together in trying to make sure we maximize the potential of the technology for good. So, we definitely welcome, be it through the GAC or other mechanisms, input from this community and to hear from all of you. On the first point about just the EU and regulation: The company has been very clear that we think that there are certain areas where regulation will be helpful. I think where the company and Microsoft starts getting more concerned is if the regulation is too prescriptive and not outcome-

focused, given the fast pace of the technology. And when we start seeing different regulatory frameworks pop up in many different countries and jurisdictions, that are sometimes divergent or sometimes even at odds with each other, because then, it quickly gets into the territory of becoming a challenge for us to navigate and implement. So, we believe that right now, we are at this cusp of really making significant breakthroughs where, to my earlier comment about the talent, we may be able to tip the scale at a systemic level in favor of the defense, so we want to make sure that we achieve that potential. I hope that answered your question.

GREG SHATAN Thanks so much, Tim, and thank you, Joanna, for the very comprehensive question. If we could turn to the Q&A pod next, and then after that, we'll go to Hadia. I mean to the Q&A pod, whoever is reading the pod.

MICHELLE DESMYTER Thanks, Greg. This is Michelle, for the record. We do have a question from [Saïd Najib]: With which method is finding AI in CleanDNS?

JEFF BEDSER Well, the method we use is in the presentation. I do know that internally, we do use Copilot for internal uses, but on the platform, I think that's probably proprietary. We are using a commercially-available, but I'm not at liberty to disclose which one.

MICHELLE DESMYTER: All right. Our next question comes from Sivasubramanian: On the other hand, can't AI be used to design phishing sites to look more authentic, to such an extent that the phishing site is free of the patterns and elements that are typical?

JEFF BEDSER: I'll take that question. So, yes, but a phishing site isn't just the landing page, and the other details about the infrastructure surrounding that site. So, for example, most phishing sites have been created for the purpose of phishing. I mean there are some compromised hosts, but the majority, it's a malicious registration to make that site. So, the age of the domain is a key factor, as well as the structure — the underlying name servers, IPs, the reputation thereof. Many times, you'll find that the underlying IP address and hosts are bulletproof hosts, and bulletproof are those — the name bulletproof comes about because they won't respond to subpoenas, or any type of process, and don't log anything. That's really not a legitimate business practice, so if you're doing that, it pretty much means you're facilitating criminal activity. So, there's a lot of factors to go with, aside from just the landing-page; but yes, the landing-pages are getting better and better using AI. There's no doubt about that.

GREG SHATAN: Thank you, Jeff. Next, I'll turn it to Hadia Elminiawi, in the room. Hadia is coming to the microphone.

HADIA ELIMINIAWI: Thank you, Greg. And this is Hadia, for the record. And my question is to Ana. So, Ana, you spoke about transparency in AI, and my question to you is: What does transparency in practical terms mean? Does this mean an AI opensource? We know, for example, that AI opensource still does not demystify how AI makes its decisions. And black-box systems remain black-boxes, even with opensource. So, in practical terms, what does this really mean? Thank you.

ANA NEVES: Thank you for that question. Very interesting. I think that when I mentioned transparency of AI, it was about transparency — accountability. So, AI models should be explainable. So, allowing DNS operators and cybersecurity professionals to understand decision-making processes; so, if you don't understand what you are fighting for, so it's very difficult. So, if AI models can be explainable, it's much easier to address what might be wrong. So, that's it, in a nutshell. Thank you.

GREG SHATAN: Thank you. Jonathan, did you have a question?

JONATHAN ZUCK: I do. I'm Jonathan Zuck, for the record. I only have my own AI, which I guess I'll call "Antique Intelligence," to bring to bear on detecting phishing attempts and things like that, but it seems like most phishing emails are tied in some way to trademarked entities or to large recognizable names. And somewhere embedded in that

email is a link to a pharming site. And I guess I'm curious as to the degree to which an email handler could look and detect who this was meant to be from and look at the links that are embedded in the email and look them up to see if they're owned by the entity that the email seems to be from. It seems like a fairly straightforward task for an AI, but I haven't seen anything like that, I guess, to date. I mean I remember being really surprised one day when I hit "send" in Gmail, and it said "hey, it looks like you meant to attach a file," and I was really excited and then a little creeped out, but you know, OK, they're reading the message. I mean on the way in, it seems like that kind of analysis says "hey, this appears to be from Norton Utilities, but the links embedded in it are Bitly, so be careful" or something like that.

TIM MAURER

I think this goes back to what we discussed at the beginning about, when we talk about AI, a lot of this is not new, and that we actually have a fair amount of systems already in place that are based on Machine Learning and algorithm, that, to your point, are the reasons why a lot of emails are going into the spam folder directly, because there are the filters that are now in place compared to 15, 20 years ago, to make sure that your inbox is cleaner than it used to be. This is where I think, to your point, and also what you raised, the real potential we hope to see with AI moving forward is to better connect some of the dots that haven't been able to connect and bring different data sources and insights together, so that you can detect if a site has been registered in a way where we have concerns about the identity because of the length of the page. And that's

where we think the real potential, maybe apart from making it more user-friendly and consumable; and to the point about explainability, figuring out what is it that the human actually needs to know and read to then act upon it. So, I hope that answers your question.

GREG SHATAN So, we're over time but I will take one last question. Amrita, please.

AMRITA CHOUDHURY Thank you so much, Greg. Amrita, for the record. I have two questions. One is, I would like to understand from basically Tim, you, and Jeffrey is: We've been speaking about English language abuse, and how AI can use or not use or aggravate it; what about the non-English speaking, because that's the global majority of what is happening there? Two is, we are seeing that this domain name abuse is growing content obviously but the — and its effect financially reputation-wise everywhere — and especially coming from global majority countries where you have new users who are not that literate in terms of digital, or even literate to that extent. So, what are you all doing for those end-users to, you know, with videos, trade tutorials, etcetera, how are you doing it? Especially with Microsoft when earlier, about 10 years earlier, you had digital literacy courses as in for people; so, for this kind of safety, what are you doing? Because people are also losing money. Thank you.

TIM MAURER

I'll start, then happy to hand it over. So, Microsoft with its business around the world is rolling out products and services and models that are in multiple languages; that goes also as far as talking about security with respect to our red teaming efforts that are not just taking place in English but also in other languages. So, we are very much committed to make sure that the technology is accessible to people around the world, and that also the safety mitigation measures and security measures are based so that they're reflective of the users that are using them. With respect to the capacity building and training piece of it, we made a big announcement last year with respect to Kenya and a specific focus on the Global South, because we are very cognizant that we want to use the technology to bring new users also online, and to make sure that they know how to use the technology. We also have a lot of internal training where we're all encouraged to use itself, so it's not just focused on specific user groups or countries; I think it's an all-up effort to make sure that we provide the training alongside rolling out new products and services so that we can maximize the impact of the tech.

JEFF BEDSER

So, thank you for that fine point, because I did neglect in my presentation to mention the fact that it was anything but English. But yes, the tools being used are done for any language that the threat comes in in, whatever the Image Analysis is, whatever the text is; and that again is one of the reasons why the AI is so effective for us, because I don't have to have someone on staff that speaks and reads every global language. The AI can do the assessment of

the language and determine what's going on and pull out the relevant components. So, it is a very big component, is making sure we handle not just the majority languages but all languages, because since we're providing services around the globe, we see phishing, malware etcetera being delivered in every language. So, we use the AI effectively to help us with the fact that I couldn't afford a staff that big that managed all those languages.

GREG SHATAN

Thank you, Jeff. We've now come to the end of our time, unfortunately; we still have some couple interesting questions. So, sorry Dana and [Atif]. But this is of course a continuing discussion. So, I'd like to thank all of our panelists — Jeff Bedser, Tim Maurer, Ana Neves — for being here. I'd like to thank you all for listening, and making this a NARALO Roundtable to remember. We'll be back in this room at the bottom of the hour for the NARALO Townhall. Thank you, and this meeting is now adjourned.

[END OF TRANSCRIPTION]