AI Training & Inference

Description: The consequences of Internet content restriction. The measured risks of third-party browser extensions. The consequences of SonicWall's unpatched 9.8 firewall severity. The incredible number of still-unencrypted email servers. Salt Typhoon finally evicted from three telecom carriers. HIPAA gets a long-needed cybersecurity upgrade. The EU standardizes on USB-C for power charging. What? Believe it or not, a CAPTCHA you solve by playing DOOM. And once we've caught up with all of that, what I learned from three weeks of study of AI.

High quality (64 kbps) mp3 audio file URL: http://media.GRC.com/sn/SN-1007.mp3
Quarter size (16 kbps) mp3 audio file URL: http://media.GRC.com/sn/sn-1007-lq.mp3

SHOW TEASE: It's time for Security Now!. Steve Gibson is here. This week a revelation. There is an incredible number of yet-unencrypted email servers out there. You don't want it to be your provider. Steve will talk about that and why it's still happening. Also a CAPTCHA that you can solve by playing DOOM. And then Steve gives us the results of three weeks of hardcore research on how AI works, a really good, I think, insight into artificial intelligence. That and more coming up next on Security Now!.

Leo Laporte: This is Security Now! with Steve Gibson, Episode 1007, recorded Tuesday, January 7th, 2025: AI Training and Inference.

It's time for Security Now!, first show of a brand new year, with this guy right here, Mr. Steve Gibson, who did not miss his Tuesday broadcast one bit; right?

Steve Gibson: You're right. As it turns out, working almost 24/7 around the clock on code can actually burn one out.

Leo: You burned out coding? I don't believe that.

Steve: I got to a point where, especially when I was - okay. So I'm working on the DNS Benchmark.

Leo: Yeah?

Steve: IPv6 has been fully supported now for a while.

Leo: Nice.

Steve: I'm now working on bringing up the TLS, the secure encrypted protocols. And the problem was...

Leo: These are all new features; right?

Steve: Yes, this is all new.

Leo: So you're had a DNS Benchmark for a long time, but you're going to do a Pro - we should fill people in who didn't hear this - a Pro version that will have additional features.

Steve: Yeah. And so here was the problem was that I wrote it 15 years ago originally. And an IPv4 address, IP address, is 32 bits.

Leo: Right.

Steve: Well, that's the size of the registers in the x86.

Leo: Oh, a little too convenient.

Steve: Yes. Yes. Throughout the entire code I'm assuming that a DNS server's IP fits in a register.

Leo: Yeah.

Steve: And so you can do so many clever things that way.

Leo: Of course.

Steve: You can index into a list using the IP address.

Leo: Ah.

Steve: You can sort the IPs by sorting 32-bit words that are the native size of the processor.

Leo: So fast.

Steve: I mean, well, and so first thing that happened was IPv6 won't fit in a register because that's 128 bits. And we want one of the big features - I've never seen any performance benchmarks about this next generation encrypted DNS, you know, DOH and DOT and DOQ, which is the QUIC protocol, the Q-U-I-C protocol, all of which this next generation of the benchmark will support. So the first thing I had to do, which is where, I don't know, the first month went - and, oh, Leo, I had to be, like, checkpointing my code. I would go try to make some changes and go down a blind alley and go, okay, well, that didn't work. So I'd restore the original source code, learning what I had learned from what didn't just work, and try again.

I mean, it was - I had to rewrite, I have had to rewrite, a huge portion of the original benchmark because it was so locked into 32 bits for an IPv4 address. And that had to be completely scrapped in order to allow both IPv6 and basically URLs because the way you address DOT, you know, DNS over TLS; DOH, DNS over HTTPS; and DOQ, you address them as URLs, not as IP addresses.

Leo: Oh, interesting.

Steve: So anyway, so of maybe...

Leo: Now you have an appreciation for what the Unix graybeards are going to have to go through between now and 2038, having represented time as a 32-bit number, which fits very conveniently to a register. They're going to have to add a few bits.

Steve: Yeah, it's - anyway. So about a month ago, I guess, IPv6, I got that all running.

Leo: Nice.

Steve: The fact that it ran at all meant that I was now - I have abstracted myself out of the IPv4 32-bit problem. That was all working. But I've never had the occasion to create a naked TLS connection because normally you just use HTTPS. And I've done that a lot on my various apps. But I've never needed to create, like to do a certificate exchange and negotiate a TLS protocol...

Leo: All that's handled underneath by the browser; right?

Steve: Exactly.

Leo: Now you've got to do it yourself.

Steve: Or a Windows API that just does it all for you.

Leo: Right.

Steve: So I had, in order to get a non-HTTP raw TLS connection, that was all new code. So that's all now in there. And I do have DOT working. Anyway, we got into all this because...

Leo: I'm impressed, actually, what you got done in a few weeks. That's very impressive.

Steve: Well, it's - yes.

Leo: And it almost killed you, didn't it.

Steve: What happened would be, after working for five days morning, afternoon, and evening, and Lorrie saying, "Honey, really you work too much," I got to a point where, if I was facing some next challenge that I had to deal with, it's like, okay, I can't do this now. I just - in the morning I'll be fresh. Anyway, what I realized was not having the weekly break, like the enforced break to switch to Security Now!, bring myself up to speed about what's been going on, read all of our listener feedback in order to, like, you know, get hints from our listeners, it actually is a good thing. So...

Leo: Yeah.

Steve: I'm glad we're back because...

Leo: Think of it as your weekend, the day and a half to two days you have to prepare for Security Now!.

Steve: Yeah. And actually that's really what it is. It is, it's a time-shifted weekend because I work on code all through the weekend.

Leo: Of course. There's no Saturday and Sunday for this man.

Steve: No. Anyway, so...

Leo: There is Monday and Tuesday, though. That's the thing.

Steve: Today's podcast, first podcast of 2025, is titled "AI Training and Inference."

Leo: Oh, I know what else you did over the break. You learned a little bit about AI, didn't you.

Steve: Yes. As I told our listeners, because I said, okay, it was going to be three weeks, right, because we had - we had the Best Of, and then we were dark on New Year's Eve,

so for me it's been three weeks since I was last focusing on the podcast. And I told everybody...

Leo: So this has to be, to be clear, what Steve has done in three weeks is figure out how to use IPv6, how to do TLS naked, and how AI works. Not much.

Steve: It was a good holiday.

Leo: Holy moly.

Steve: So before we launch into the podcast, I want to take a moment to assure everyone who's like, oh, god, not more AI, that this podcast which we call "Security Now!" is not morphing into "AI Now!"

Leo: Good.

Steve: I'm quite conscious of the fact that through the end of 2024, and yes, here today, you know, we have and will spend time looking at what's been quietly simmering in the back rooms of university and commercial labs for years and has just suddenly, you know, burst out onto everyone's foreground attention. You know, and of course, you know, historically from time to time we've veered rather far afield, touching on topics of health, science fiction, the Voyager spacecraft, and even homemade portable sound guns. What underpins all these diversions is the underlying science and technology that makes them go. And in this most recent case, you know, my focus and fascination with AI, you know, all of the feedback that I've received from our listeners has suggested that this is a topic of interest...

Leo: Oh, yeah.

Steve: ...that is deeply shared. And in fact we've got a bunch of listeners who are in AI. We've got Google AI listeners among those here. So, you know, over the holidays, during the three weeks we've been apart, as we said, I focused upon bringing myself up to speed, really, about what's been going on. And I've come away with an understanding, I think, of the big picture. And I have a number of observations that I'm excited to share. So we'll get to that.

But I also think that this is probably it for a while. I'm sure that eventually the fallout from AI research will bear directly upon the security of our software. I don't know how, you know, Microsoft must have a team because, you know, they're sharing in a lot of the OpenAI technology, being a major investor. They must have a team - I hope they do - who are already thinking, how can we leverage this to have fewer patches on every second Tuesday of the month. So anyway, I wanted to assure everyone, yes, we're going to talk about it again at the end of today's podcast. But not forever. I really think this gets it out of my system, and I will be now content to wait for things to mature.

But we're going to talk about more than that, of course. We've got - we're going to talk about the consequences of Internet content restriction. The measured risks of third-party browser extensions. There have been some more troubles there. The consequences of SonicWall's unpatched 9.8 seriousness, you know, CVSS score firewall severity. The

incredible number of still-unencrypted email servers, Leo, meaning not individual email encryption, but the interchange of email among servers still not encrypted today.

Leo: That's a shock. People are sending their passwords in cleartext, in other words.

Steve: Just wait, yes, yes, exactly.

Leo: Wow.

Steve: And the content of their email. I mean, everything is in the clear.

Leo: That's shocking.

Steve: Also, and I heard you mention this, I think it was on Sunday, we have the declaration, we hope it's true, that Salt Typhoon was finally evicted from three telecom carriers. They've all said, you know, Verizon...

Leo: So they say.

Steve: Oh, yeah, they're all gone now. Yeah, right.

Leo: So they say.

Steve: Uh-huh. Also HIPAA is getting a long-needed cybersecurity upgrade. The EU, oddly, has decided to standardize on USB-C for its power charging.

Leo: Yeah?

Steve: What? And then, believe it or not, we have a CAPTCHA you solve by playing DOOM.

Leo: Wow, that's funny.

Steve: And once we've caught up with all that, I'm going to share what I've learned from three weeks of studying AI technology. And of course we have also, as our Picture of the Week, Security Now!'s first-ever caption contest. So...

Leo: Well, this will be fun.

Steve: It's going to be fun.

Leo: And those of you watching live, don't look. Hold your powder. We'll give you a chance, too, to caption the upcoming Picture of the Week in just a moment. It's going to be a good show. Okay. Caption contest time, Steve. Do you want to prepare us in any way for this?

Steve: Well, so you can just look at the picture.

Leo: Okay.

Steve: And it raises more questions than it answers.

Leo: Yeah, what's it protecting would be question number one.

Steve: Yeah. And what I love is that you can sort of see a bit of a path, out from the vantage point of the photographer of this, to the gate. So for those who can't see, it's just this bizarre - normally you can sort of figure out, okay, what one of these strange pictures, how it came to pass. We have a metal security gate with bars and a locking plate that's protected so you can't slip a credit card in, and a locking handle - out in the middle of a field.

Leo: This is the field that Steve says you have to go to to have completely private conversations.

Steve: Exactly.

Leo: Maybe that's what it's protecting. I don't know.

Steve: It hasn't been mowed for a decade. We've got, you know, bushy trees in the background. Someone said looks like - one of the plants behind it looks like a cauliflower something. Okay. But it's like, what, I mean, how do you explain this? I just - it's crazy. So as I was looking at this thinking this is a crazy photo that would be great for the podcast, and coming up short for a caption that I loved, I thought, okay, let's leave this to our listeners.

Leo: I love it.

Steve: Let's turn this over to everyone who sees these every week and gets a kick out of them. So anyway, this is Security Now!'s first caption contest. Here's the picture. It's in the show notes. Take a look at it. You know, you can write to securitynow@grc.com. I sent the email, the show notes and so forth, out to all of the subscribers to that list last night. And I forgot about the caption contest as being a thing. And I thought, what is all this email coming in? Like, immediately.

And that's why, before the podcast, I asked you, Leo, I think you're going to have to explain to me what's going on with Narnia because, if there's one term I've heard more than any others, I mean, we've had I should say already a bunch of great submissions.

Don't let that forestall anybody from sending theirs in. Next week we will have the, what, the top 100 captions that have been suggested out of the thousand that I imagine that I'm going to be receiving.

Leo: And now you know what Narnia is, of course, it's a magical kingdom from the book "The Lion, the Witch, and the Wardrobe." And you get to it by going through the back of a giant wardrobe closet.

Steve: Yes. And this does look like maybe...

Leo: You're going to Narnia.

Steve: You can't tell from looking at this, this is actually a portal to somewhere else. Because it looks like you're actually seeing this...

Leo: That makes sense, actually.

Steve: ...this shrubbery behind the gate. But no, if you - and clearly some people have walked down that path from her to the gate, probably just to check, you know, jiggle the handle and see if the gate's locked or not.

Leo: It's an attractive nuisance, for sure.

Steve: Yeah.

Leo: We're getting some suggestions from the chatroom, like, "Oh, I forgot my key" would be one. And "The long-forgotten protocol" is another. But I bet you the best way to do it would be to email Steve. Is there a prize for the best caption?

Steve: No. Hearing yours read out loud...

Leo: On the show.

Steve: ...on the podcast.

Leo: Yes, there you go.

Steve: They'll be, like, that was mine.

Leo: That's your prize.

Steve: That's the one I sent.

Leo: That's your prize. Awesome. All right. Well, let's get going. We've got a show to do here.

Steve: We do indeed.

Leo: You've got lots of stuff probably happened in the last three weeks.

Steve: Okay. So I know you touched on this a little bit on Sunday, sort of tangentially. But questions surrounding restrictions on access to Internet content are both controversial and nuanced. You know, they factor in the individual's age and their location, the nature of the content, and the prevailing government. And, you know, if 10 different people are asked about restrictions on access to Internet content, you're going to get 10 different answers back. So not a lot of consensus there. And where questions of access to Internet content by children arise, even parents and guardians will disagree.

But I do know from conversations with many parents of young children, many of whom take time from their lives every week for this podcast, managing what their kids are exposed to on the Internet is a source of significant concern. The first thing many of our listeners do when setting up a new network at home is to choose a DNS filtering provider that offers what's known as a family-oriented plan which filters out and removes access to the Internet's more unseemly websites.

Now, one place where everyone, I would say nearly everyone agrees is that "age appropriateness" is a thing. You know, there's content on the Internet that requires some maturity and perspective to understand correctly. Back in the days before the Internet, you know, which is a world that many of us remember well, our rough age could be determined just by a glance at us; right? So if at the tender age of 10 or 11 we were to try to get into a bar or a strip club, those who stood to lose their license to operate such a facility would go to great lengths to prevent our entrance. And, you know, everyone's familiar with the concept of a fake ID. The only reason of needing to fake an identity is to enable its holder to do something that the law forbids them to do at their true age.

But what's different today is that we have the Internet, and no one knows how old anyone is in cyberspace. Although there can be benefits to this, it's also subject to abuse. And this represents a profound change from the physical world that many of us grew up in. Having been born in '55, I was 34 years old by the time that in 1989 Tim Berners-Lee came up with the idea for the World Wide Web. That means that there was never a time for me when a website might ask me to verify that I was at least 18 years old, and that wasn't true. You know, I was nearly twice that age by the time that websites started thinking that would be a good thing.

But there's no doubt that with gossip and curiosity and peer pressure being what it is, plenty of today's children who are probably far short of their 18th birthday might well be clicking those "You betcha I'm 18!" buttons. It's not my intention to moralize, and I'm not doing that here. If today's Internet existed when I was 14, I have no doubt that I would have been curious to see what was hidden behind those buttons and that I might have been pressing them after first bouncing my connection through a handful of Tor nodes.

Now, I suspect that few parents would disagree that where age appropriateness is concerned, a world of difference separates access to the sort of hardcore adult content

that's readily available on the Internet from viewing TikTok cat videos. And the difference is so stark that the Internet's premiere adult-content website already blocks its access across much of the U.S. Southern states, and it just went dark across all of Florida last Wednesday, in a preemptive action as the Sunshine State's latest legislation went into effect. A lot of this legislation happened here at the beginning of 2025.

Okay. So that's on the extreme side. But what about the cat videos? I chose this as our first topic of 2025 because, as we start into this new year, as I said, more and more states are enacting and have enacted Internet age restriction legislation aimed at the far more benign gray area of modern social media. And much of this new legislation that just went into effect at the beginning of the year is ad hoc. You know, I think because we've been addressing the issues for a while, it's increasingly well understood that there are pros and cons to this. But if you look across the legislation, it's just random and uncoordinated.

Here's a really brief timeline. On July 1st, so summer before last, 2023, Connecticut put legislation called SB 3 into effect which requires social media platforms to obtain parental consent before allowing minors to open accounts. Then jump forward a year to last summer. On July 1st of last year, Louisiana's Act 456 requires social media platforms to impose limitations and restrictions on certain accounts, implement age verification for account holders, and obtain parental consent. A couple months later, September 1st, that's four months ago, Texas HB 18 requires digital service providers such as social media platforms to get consent from a parent or guardian before entering into an agreement with minors younger than 18, including to create an account.

On the 1st of October, Maryland Kids Code, as it's called, requires social media platforms to set default high privacy settings for users under 16, ban the collection of children's data for personalized content, ensure age-appropriate design, implement age verification, and obtain parental consent for younger users. The same month, Utah HB 464 and SB 194, you know, House and Senate in Utah respectively, the Social Media Regulation Act requires parental consent for minors to create social media accounts and mandates age verification by social media companies. It also restricts social media use between, okay, 10:30 p.m. and 6:30 a.m. for users under 18 without parental consent. Okay.

First of January, so 2025, Tennessee HB 1891 requires social media companies to verify the age of users attempting to create and maintain accounts. It mandates that platforms obtain parental consent for minors under 18 and enforces stricter privacy and safety measures for these users. The law aims to protect minors from potential online harms by ensuring that social media companies comply with these new regulations. There were also three others that passed and will be coming into effect. Florida, the one I mentioned before, HB 3, requiring social media platforms to verify users' ages, obtain parental consent for users under 18, protect minors' personal data, limit their exposure to harmful content. Georgia's SB 351, known as the Protecting Georgia's Children on Social Media Act of 2024, requires social media platforms to implement age verification processes for users, mandates parental consent for minors to create accounts, and restricts social media use in schools.

And finally, Minnesota MN HF3488 sets rules for compensating minors who contribute to online content creation. What? You're going to compensate them? It requires content creators to keep records and set aside earnings for minors, and it allows for legal action against violators, also mandates the removal of content featuring minors upon request. And I should mention also, I didn't put it in the show notes, but the penalty in Florida is $50,000 per infraction.

Leo: Per minor.

Steve: Yes.

Leo: Yeah.

Steve: It's like, what? Okay. And on top of all this our U.S. Congress also has some legislation that's been floating around since 2023 known as the Protecting Kids on Social Media Act, and its future's unclear. And I have no idea what position the incoming administration and our next Congress will adopt on such measures. You know, on the one hand there's the politically popular promise of "protecting the children," whereas the flipside is that pesky the U.S. Constitution's First Amendment guarantee of freedom of speech. And I should mention that a bunch of this new legislation is already under injunction because First Amendment says you can't do some of these things, legislators, no matter how much you want to.

Now, a well-known website featuring adult content greets its visitors with this statement. It says: "Did you know that your government wants you to give your driver's license before you can access this site?" It says: "As crazy as it sounds, it's true. You'll be required to prove you are 18 years or older such as by uploading your government ID for every adult content website you'd like to access. We don't want minors accessing our site and think preventing that from happening is a good thing. But putting everybody's privacy at risk won't achieve that."

Now, of course it's unclear what would prevent anyone from uploading a photo of someone else's ID, or just synthesizing one from scratch to upload. You can imagine a bunch of websites will pop up, you know, the Create Your Own ID site. But the larger point here to note is that there are consequences to this move from the real world to the cyber world, and that the unfettered anonymity and freedom we've enjoyed through the first 24 years of the 21st-century Internet may soon be challenged.

Now, it may be that none of this will come to pass, or that, at least if it does, it won't be until its consequences have received significant legal and constitutional scrutiny. In reaction to Florida's new laws, last October the Computer and Communications Industry Association and NetChoice, whose members include the likes of Google and Meta, big social media platform providers, filed a federal lawsuit challenging the constitutionality of the various restrictions being imposed by this new Florida law. The lawsuit's text stated: "In a nation that values the First Amendment, the preferred response is to let parents decide what speech and mediums their minor children may access, including by utilizing the many available tools to monitor their activities on the Internet."

Now, this feels as though it's headed to the Supreme Court because U.S. legislators are going to need to have some clarification about what they can and cannot require of social media and other companies. But what seems clear today is that these long simmering issues are beginning to come to a boil, and that the parents and guardians of minors may soon be put in the loop, at least, and given the controls hopefully which they need to allow their households to abide by whatever the prevailing laws end up being for their locality. But the question is, how can this also be done while preserving the privacy of the individual? As I started out saying, no one knows how old anyone is in cyberspace. That also applies to you and me; right?

No one looking at me today in the physical world would mistake me for a minor. But when any of us connect to any website, there's no indication of any kind how long we've been breathing this planet's air. There's been a freedom that we've all enjoyed up to now. So we need to consider what it means to have that change, since that's what we're talking about here. No one would argue that our children need to be protected from

harm, even while we're going to need to work out an exact enough definition of harm to be actionable. And that's going to be a challenge. But as that notice on that premiere adult content website noted, the ultimate consequence of that may be us needing to somehow affirmatively show that we're not minors who are in need of state-mandated protection. How do we do that without sacrificing a great deal of the privacy we currently enjoy? I don't know, Leo.

Leo: Yeah. As you know, we talk about it a lot on all of our shows. Australia passed a law banning all social media for kids under 16.

Steve: Right, like a few months ago, and we did talk about that.

Leo: It's not in effect. It won't be in effect till the end of the year. But their attitude is, well, we don't know how to do this. But you guys are smart. You figure it out.

Steve: Well, and we saw how well that worked for the encryption problem; right?

Leo: Yeah.

Steve: It's like, we need to be able to see what people are doing, and we don't know how. So you guys are smart. You guys, you know, you techies, you just figure out how to give us what we want and not breach anyone's privacy. No, I really - the biggest point I wanted to sort of point out here is that the physical world figured out how to do this a long time ago, and that's the world we grew up in. But in cyberspace it really, I mean, it's easy to forget that anonymity is something that we sort of take for granted with our use of the Internet. But that's at odds with exactly what all of this legislation which we're now seeing begin to happen wants to do. It says, you know, we need to know how old you are. And that's a huge change. And it's not just how old children are. They need to know how old we are to know we're not children.

Leo: Yeah, I got carded the other day, and I thought, that's hysterical. But the guy said, well, it's policy. We know obviously you're not under 18 or under 21.

Steve: I was, too. I was trying to remember where it was. Somebody asked for my ID. I said, what?

Leo: This was at a Cost Plus, one of those import stores. And he just said, yeah, we just do it. I said, "I'm not even buying the liquor. This old lady is." And he said, "I need hers, too." There is a cynical side of me that says, and this is true I would say in Texas, Louisiana, a few states, where they don't want this to be solved. They want to ban pornography. And so they don't really care if this can't be solved. They're happy. And it's happened in a number of these states, including just now in Florida, where these big pornography sites just abandon the site, abandon the state, say, well, you can't use this.

Steve: They can't afford the lawsuits. It's just not worth it.

Leo: And I think honestly that's what the legislators want. Seriously, that's what they're trying to do is ban pornography.

Steve: Is to scare the adult websites out of their state.

Leo: Yeah. They don't like pornography. That's a whole different argument, and it doesn't have a security angle to it. But, you know, we live in interesting times, don't we.

Steve: Well, and for me, we've talked about this a little bit, and yes we do live in interesting times, which is why I'm so glad we're here now, Leo.

Leo: No kidding.

Steve: And you and I are talking about this.

Leo: Especially, by the way, for AI, because that's about to change everything in ways that may make this trivial; right?

Steve: So for me, the question is the technology of this; right? Because we've talked about the technology of tracking. We've talked about the technology of encryption. Well, what about the technology of age attestation? Like how do you do that? Because one of the things that upset us about that first Google attempt at eliminating tracking was where, when you visited a website, it would present that token that told the site about your interests. And everyone said, and I remember you saying, you know, quite rightly, wait a minute. You know? They don't have that now. So suddenly our web browser is going to be telling every site we visit what our collection of interests are.

Leo: Hey, Leo's really interested. You got any pornography? Yeah. These are such difficult problems. I just read a statistic, and I think it's probably accurate, that said, in order to change a policy, any policy in this country, it takes 90% of the people to believe it should be changed. Not 50%, not 60%, 90%. There has to be a generally obvious consensus.

Steve: An overwhelming...

Leo: An overwhelming consensus that this is what we should do. And that happens so rarely on any subject that it seems nothing much happens ever. I don't know. It's a quite interesting issue, and...

Steve: One that we are going to be facing. We, you know...

Leo: Paris Martino did a very interesting piece in the Information Weekend about a new kind of a face recognition technology, I think it was called Yoti, Y-O-T-I, that did age verification. And so that's what I think legislators and companies are looking for

is something passive, that it just looks at you, you don't even have to pose, it just says, yeah, you know, you're probably over 16; or, no, you're probably under 16. I mean, maybe that's the solution? The people at Yoti claim it works quite well.

Steve: Of course it does mean that you have to have a camera aimed at you.

Leo: Oh, that's a good point. Yeah, many people probably don't want that either.

Steve: Yeah, it's a little spooky, you know, yeah. What's not spooky is this next advertiser.

Leo: Oh, they're fantastic. In fact, your timing couldn't be better, Steve. Because you know what happened when those laws passed in those states? VPN sales went through the roof.

Steve: Uh-huh.

Leo: Yup. And guess what? A VPN protects your privacy. Every sponsor you hear on this show and our other shows in the new year, they've re-upped, and we're very grateful to them. We're also grateful to all the brand new subscribers we got. You know, I made the pitch in the last few weeks of the year that we may not make it in 2025 without your help, and a lot of people have joined Club TWiT thanks to that. So welcome to our new Club TWiT members. And as always an invitation to everybody to join if you're not a member: TWiT.tv/clubtwit.

All right. Let's go on. Sorry to interrupt for such a long period of time. Back to Mr. Gibson.

Steve: So we have a bit of a cautionary tale here.

Leo: I think everything on this show is a cautionary tale, to be honest.

Steve: That's true. Except AI. I don't think that's cautionary, at least not...

Leo: Well, I'll be interested in what you have to say, actually. I'm very curious, yes.

Steve: We'll see. Okay. So I needed to share this because it highlights a very real threat which users of increasingly popular web browser extensions face. And that's a compromise of the extension, which is then downloaded or updated by the user's browser. Now, several times in the past we've talked about the threat of an extension's author abandoning an extension, like deliberately saying, "Okay, I'm done with this, I've been tending this thing for 10 years," and then selling his, you know, basically the install base to an unscrupulous third party. So that's one problem.

But there's a different one. The other clear and present danger is a deliberate attack on and compromise of an extension's publisher for the purpose of turning an extension

malicious. This is what recently happened to the cyber firm Cyberhaven, the security firm Cyberhaven, and at least 35 other known Chrome browser extensions that are known to have been compromised as part of a concerted effort. Okay, so what happened? Two days after this past Christmas, on December 27th, Cyberhaven posted under their headline "Cyberhaven's Chrome Extension Security Incident and What We're Doing About It."

Leo: You do not want that headline. Oy. Oy oy oy.

Steve: They wrote: "Our team has confirmed a malicious cyberattack that occurred on Christmas Eve, affecting Cyberhaven's Chrome extension. Public reports suggest this attack was part of a wider campaign to target Chrome extension developers across a wide range of companies. We want to share the full details of the incident and steps we're taking to protect our customers and mitigate any damage. I'm proud," writes the author of this, "of how quickly our team reacted, with virtually everyone in the company interrupting their holiday plans to serve our customers..."

Leo: Oh, that's why they do it Christmas Eve, isn't it.

Steve: That's exactly right.

Leo: Nobody will be home.

Steve: That timing was no coincidence, "...and acting with the transparency that is core to our company values." And I've got to say, and I will say, I'm impressed by this response. The guy wrote: "On December 24th, a phishing attack compromised a Cyberhaven employee's access to the Google Chrome Web Store. The attacker used this access to publish a malicious version of our Chrome extension, which was version 24.10.4. Our security team detected this compromise at 11:54 p.m. UTC on December 25th and removed the malicious package within 60 [six zero] minutes."

So they have some bullet points. "First, version 24.10.4 of our Chrome extension was affected. The malicious code was active between 1:32 a.m. UTC on December 25th and 2:50 a.m. UTC on December 26th, so for a total of a little over 25 hours. Chrome-based browsers that auto-updated during this period were impacted. Our investigation has confirmed that no other Cyberhaven systems, including our CI/CD process and code signing keys, were compromised. For browsers running the compromised extension during this period, the malicious code could have exfiltrated cookies and authenticated sessions for certain targeted websites." Now, they know that it's Facebook.com. We'll get to that in a second. Also, "While the investigation was ongoing, our initial findings show the attacker was targeting logins to specific social media advertising and AI platforms.

"Then our response: We notified affected customers December 26th at 10:09 a.m. UTC. We also notified all other customers not impacted. The compromised extension has been removed from the Chrome Web Store. A secure version, 24.10.5, has been published and automatically deployed. We have engaged an external incident response firm for third-party forensic analysis. We are actively cooperating with federal law enforcement. We've implemented additional security measures to prevent similar incidents.

"For customers running version 24.10.4" - that's the bad one - "of our Chrome extension during the affected period, we strongly recommend: Confirm if you have any browsers

running the Cyberhaven Chrome extension version 24.10.4 and force an update to version 24.10.5," they said, "currently available in the Chrome Web Store, or newer. Rotate Facebook personal and business account passwords for accounts on impacted machines. Review all logs to verify no outbound connections to the attacker's domain or other malicious activity."

Okay. So it's good to see that this security firm acted appropriately in every way. They responded immediately. They determined the original attack vector, how the bad guys penetrated their perimeter security, and they now know that an employee fell victim to a crafted phishing attack. They replaced their compromised extension quickly, verified that this was the extent of the penetration, and notified the public without delay. They fessed up to the mistake and made no attempt to downplay it. And they did all this on Christmas Day.

Leo: Wow.

Steve: So as you said, Leo, it's likely no coincidence that the phishing email attack was launched on December 24th, the day before a span of holiday that was doubtless intended to maximize the period of time the extension's malicious modification would go undetected.

Now, I'd have to say that this particular phishing attack might have caught any developer unaware. The show notes here, adjacent to the text here on page six, has a snapshot of the perfectly formatted HTML notification that was received by a developer. I mean, it looks completely legitimate. You know, from the Chrome Web Store: "Hi there. We wanted to let you know that your item is at risk of being removed from the Chrome Web Store. Please see the details below." Then it gives it the item name, Cyberhaven security extension v3; the item ID, which is correct. And then under Violations it says: "Excessive and/or irrelevant keywords in the product description." Which, you know, okay, whoops.

Leo: It happens, sure.

Steve: "Violation: Unnecessary details in the description." And then it says "Relevant section of the program policy." And then it quotes their policy that somebody felt at Google or Chrome Web Store management was wrong. And then there's a button for Go to Policy.

Leo: Yeah.

Steve: So, I mean...

Leo: Who wouldn't click that?

Steve: It looks like a completely legitimate event. Once the employee clicked on the email, they were taken to the standard Google authorization flow for adding a malicious OAuth Google application which was called, and it shows it on the screen, "Privacy Policy Extension." Which if you really stop to think about it, it's like, whoa, wait. I'm authorizing the addition of something called Privacy Policy Extension. Well, they named it that in order to be tricky because that's not something you want to do. But by naming it Privacy

Policy Extension, you sort of obscure that fact. So again, you know, on Christmas Eve it's like time to go home, but we don't want to, you know, we don't want to have our extension yanked during the holidays, so let's take care of this now.

The authorization page was hosted on Google.com and was part of the standard authorization flow for granting access to third-party Google applications. So just one tiny little glitch in an otherwise normal authorization flow. The employee followed the standard flow and inadvertently authorized this malicious third-party app. The employee had Google's Advanced Protection enabled and had multifactor authentication covering the account. The credentials were not compromised. Yet this still happened. So it was a very carefully crafted phishing attack designed to capture even somebody who was paying attention.

So what they found was that the malicious extension 24.10.4 was based on a clean previous version of the official Cyberhaven Chrome extension. So the attackers went to some effort in order to create this attack to set this up, and not just for them. And remember I said 30-some other extensions were all compromised. The attacker made a copy of the clean extension, then added their malicious code to create a new malicious version of that 24.10.4, then uploaded it to the Chrome Web Store. The Cyberhaven guys reverse-engineered the malicious modification to their extension in order to determine what it was doing.

In a subsequent posting they wrote: "In our analysis of compromised machines, the extension was targeting Facebook.com users. If the user was logged into Facebook and navigated to the Facebook website, the extension would execute the malicious code path. Here is what the malicious flow would execute. It would get the user's Facebook access token," meaning an impersonation attack immediately. Anybody who had that could just open their browser as them and be logged in just as they are. "Get the Facebook user's ID. Get the user's account information via the Facebook API. Get the user's business accounts via the Facebook API. Retrieve the user's ad account information, again through the Facebook API. Package all this information, along with Facebook cookies and the user's agent string, and send it to their command-and-control server."

They said: "After successfully sending all the data to the command-and-control server, the Facebook user ID is saved to browser storage. That user ID is then used in mouse-click events to help the attackers with two-factor authentication on their side if that's needed." So again, a high-level attack against browser extensions.

So the web browser extension attackers were interested in attacking the accounts of any Facebook users whose Chrome browsers might update to the malicious extension before it was detected and removed from the Chrome Web Store. Obtaining a user's Facebook access token cookie, as I said, allows full impersonation of the user. And, because Facebook now has a very feature-complete API, a lot of damage can be done.

Another security site, Secure Annex, provided a broader perspective - because, you know, the Cyberhaven guys were just focused on theirs, but this was, as I said, a much broader attack. Secure Annex provided that perspective into the attackers behind this campaign. By pivoting from the known-malicious Cyberhaven extension, indications of compromise were obtained. That's how we know now how many more Chrome web extension developers fell victim to these phishing attacks. The earliest known instance of one of this group's many attacks was way back last May. So these guys have been active since then.

I think it's important for everyone to have some sense for the scope of this. So here's, for example, 19 of the compromised Chrome web extensions: VPNCity with 10,000 users; Parrot Talks with 40,000 users; Uvoice with 40,000 users; Internxt VPN with 10,000 users; Bookmark Favicon Changer with 40,000 users; Castorus with 50,000;

Wayin AI with 40,000; Search Copilot AI Assistant for Chrome with 20,000; VidHelper Video Downloader with 20,000; AI Assistant, ChatGPT, and Gemini for Chrome with 4,000; Vidnoz Flex video recorder and video share with 6,000; TinaMind, the GPT-4o-power AI Assistant!, with 40,000; Bard AI chat with 100,000 users; Reader Mode with 300,000 users; Primus, which was previously PADO, with 40,000; GPT 4 Summary with OpenAI, 10,000 users; GraphQL Network Inspector with 80,000 users; YesCaptcha assistant with 200,000 users; and Proxy SwitchyOmega with 10,000.

So every one of those Chrome web extensions was compromised last year, and there are more. Just those exposed as many as 1,060,000 users of Chrome to malicious browser-side code. Now, the good news here, if there is any, is that the attackers appeared to be focused solely upon Facebook users and their accounts. But that was this time, and they are certainly willing, obviously, to go well out of their way to compromise those accounts.

It wasn't long ago that we were talking about the move from Chrome's v2 extension manifest to the significantly more limited v3; and how, as a consequence, uBlock Origin, for example, the full uBlock Origin, won't ever be offering its full-strength v2 version under v3, once Chrome completes that switch. I'm certain that the Chromium team understands how much value the third-party browser extension ecosystem brings to their Chrome browser. But given this attack campaign as just one example, and you've got to know they know way more about abuse of this than is even publicly known, it's not difficult to see why they would be anxious to curtail the damage that aberrant extensions are able to do to those extensions' users. Thus the move to the more limited scope v3 manifest.

And note that none of this is ever about an extension's user doing anything wrong. That never happened. It was the extension's developers whose account was accessed and abused. So this is another form of supply-chain attack. As users of Chrome, the one thing we can do is practice good what I would call "browser extension hygiene," meaning keeping the set of extensions which we're loading and using to a minimum and removing any "dead wood" that might needlessly expose us through that extension's inadvertent compromise. Every additional extension that is loaded has access to deep user data in the browser. So there's nothing you can do to prevent the extension from being compromised, but so just minimize the number that you're using. And, you know, when you look at that list, there's a bunch of crap there.

Leo: It's all crap. A lot of the stuff was AI assistants to work with the AI that you don't need.

Steve: Right.

Leo: However, just it's clear with this very effective phishing attack that it doesn't have to be crapware. It could be anything; right? I mean...

Steve: Yes.

Leo: Is there something about browser extensions that are inherently insecure? I know, I remember Google saying, oh, you shouldn't use browser extensions for your password manager because they're inherently insecure, because this was a bid to get you to use Chrome's password manager.

Steve: Well, consider that when we enter a username and password, our password manager pops up and says, would you like me to save that for you?

Leo: Yeah, yeah.

Steve: It has, it sees our username and password.

Leo: It has permissions, yeah, yeah, yeah. It has a lot of information.

Steve: Oh, goodness. Yeah. I mean...

Leo: And they're all written in JavaScript. Is that inherently problematic? Or not really?

Steve: No, it's possible to write - no. In fact, here the extensions are not the problem; right? It's that somebody crawled into the...

Leo: Yeah, they've been socially engineered, yeah, yeah.

Steve: Exactly. Well, they crawled into the developer and turned the extension malicious.

Leo: Right.

Steve: Added deliberate code to the extension, and then rode the developer's coattails, you know, uploaded an update to the extension, just like the developer would if they were fixing a bug in their extension.

Leo: Yeah.

Steve: And then of course Chrome wants to remove any bugs that might be in extensions, so it's checking to see if there's a new version; and, if so, get you the new one.

Leo: So is there an argument for not using any extensions at all?

Steve: There's an argument for it, but that would cripple us. I would, I mean, you know, we want Bitwarden to be able to auto-populate our login fields.

Leo: Sure. I do like what Brave has done in response to...

Steve: And we want uBlock Origin.

Leo: ...Manifest v3 because that will eventually turn off uBlock Origin. Brave just built it into the browser. So maybe that's the better way to do it. If it's a browser company you trust, let them handle the password manager and all of that.

Steve: Well, yes. And that's - you bring up a good point, which is you are trusting the security provisions of every extension developer whose extension you load. You know, you can imagine the lengths that the Chrome team go to to make sure that the base browser is secure. And even then there's the occasional error.

Leo: All the time.

Steve: Yeah.

Leo: And really the reason is these browsers are your interface to the outside world. So there's [crosstalk] vector. Yeah.

Steve: It's an OS now.

Leo: And it's an operating system, yeah. It's a very complex piece of software.

Steve: It's become so - as I said a long time ago, it's no longer possible to create one from scratch. You can't.

Leo: Yeah, right.

Steve: You don't have to now because Chromium core is open source.

Leo: You can use - yeah, right.

Steve: So you don't have to. But, yeah.

Leo: Yeah. I mean, I use - I'm looking at my browser extensions. I use a Chrome-compatible browser called Arc. I've got Bitwarden. I've got Snowflake. I didn't put that on there. Let me take that off. I've got uBlock Origin. Those are the two I have to have pretty much everywhere.

Steve: Yes. I would say your password manager and uBlock Origin, two must-have tools.

Leo: Oh, I know what Snowflake is. That's the thing we recommended that enables Tor to work in...

Steve: Oh, right, right, right.

Leo: Yeah. I'll leave that. I forgot about that.

Steve: Yup.

Leo: Yeah. I'll turn everything else off, though.

Steve: Okay. So Leo, we're an hour in. Let's take a break, and then we're going to get to SonicWall and some more news from the last three weeks.

Leo: Yay. Loving the news. Loving it all. And just a reminder, Steve, we're going to have an extra break in the show.

Steve: I've already - that's the pace we're keeping.

Leo: Yeah. We're very happy about it, actually. All right, back to Steve.

Steve: Okay. So back in August, SonicWall, a well-known manufacturer of popular Network Security Appliances - and now NSA has got two meanings. It's the National Security Administration, is that anything?

Leo: You know, it's funny, I should know that. We must be getting old, Steve.

Steve: I think we are.

Leo: National Security Administration. I believe that's correct, yes.

Steve: Okay. Also Network Security Appliances. NSA, Network Security Appliances.

Leo: Oh, okay.

Steve: Anyway, SonicWall revealed a serious vulnerability in their SSL VPN Firewall product.

Leo: Uh-oh.

Steve: Now, they rated it with a severity of 9.3. However, NIST officially gave it a 9.8, which, you know, that's not good. And shortly afterward CISA formally warned of the serious potential for its exploitation. They, both CISA and SonicWall, they called it the

SonicOS, which is the OS in their appliance, "Improper Access Control Vulnerability," which already doesn't sound good, and noted that it was "potentially," in quotes, being - well, they didn't have it in quotes, but everybody else has - being successfully attacked in the wild.

Now, among the reporting on this, I particularly liked the write-up by the security intelligence firm Field Effect. They wrote: "While it's unclear what SonicWall means by 'potentially' exploited, Field Effect can confirm that we have seen an increased targeting of SonicWall firewalls since CVE-2024-40766 was announced on August 23rd. However, further investigation is required to determine if the threat actors are specifically targeting 40766 or other, older, unpatched vulnerabilities." I really thought this was interesting. They said: "Traditionally, when vendors disclose critical vulnerabilities in edge devices, it draws attention of threat actors toward the devices in general, and that could be what we've observed in relation to the SonicWall firewalls." So I really appreciated their measured response. There's no breathless hyperbole here.

They finished by noting: "SonicWall firewalls are very popular among critical infrastructure industries and corporate environments and are thus frequently targeted by threat actors looking to obtain initial access into networks of interest. According to the Shadowserver Foundation" - and you're going to be hearing about Shadowserver Foundation a couple more times before we're done here today. They said: "Approximately 400,000 SonicWalls are deployed worldwide, representing a significant potential attack surface for threat actors who possess SonicWall exploits."

Okay. So that was back in August, where and when we have an estimated 400,000 Internet-facing SonicWalls with a known remote authentication vulnerability. This was three generations. Generation 5, 6, and 7 all had this vulnerability. So here we are now. Where are we? Two days after Christmas, on December 27th, a Japanese security researcher posted his own update on the state of play with SonicWall devices today.

He wrote: "In August 2024, the SonicWall NSA vulnerability 40766 was disclosed." He said: "I have found strong indications that the ransomware groups Akira and Fog are still exploiting this vulnerability for unauthorized access. Through my ongoing investigations, I found that, as of December 23rd, 2024, the number of companies suspected to have been compromised by these two groups via this vulnerability had exceeded 100." Okay. So, you know, here we're on the edge of the corporate network facing the Internet. Oftentimes we're just talking about oh, look, they got hit by ransomware. How did that happen? Well, this is how that happens. Here this guy has identified these two ransomware groups, Akira and Fog, that have used this vulnerability which was announced and for which a patch was available last August, having penetrated 100 companies that did not patch.

He says: "In this article, I will share the details of this investigation and highlight the current situation in which at least 48,933 devices remain vulnerable to CVE-2024-40766." In other words, that was August a patch was made available and announced. Today, 48,933 of those devices are still vulnerable. And in this case these two groups are known to have gotten into a hundred organizations that didn't bother to update their SonicWall.

He said: "Since the vulnerability was disclosed, I have been investigating whether the organizations listed on various ransomware groups' leak sites own SonicWall Network Security Appliance devices. Focusing on the 218 organizations identified as victims of Akira and Fog, I found that over 100, approximately 46%, were running SonicWall. Considering that the SonicWall network security appliance ownership rate among organizations victimized by other ransomware groups, excluding Akira and Fog, remains around 5% or less, this figure of 46% for those two groups is remarkably high."

In other words - me speaking - whereas the general rate of overall SonicWall presence among companies who have been breached and listed by ransomware groups other than Akira and Fog is down at 5% - still not great, but we can't blame SonicWall for like being the cause - the fact that around 46% of the organizations victimized by just those two ransomware groups, which are currently exposing a SonicWall device to the Internet, strongly suggests that those two groups have successfully designed an exploit for the vulnerability and are working their way through the inventory of still-exploitable and unpatched SonicWall device owners.

This Japanese researcher wrote: "I developed a proprietary method to evaluate patch status by examining the HTML structure of SonicWall devices to assess mitigation efforts for the CVE-2024-40766." Now, I'll just stop right there and say the fact that you're getting HTML from a device exposed to the Internet, you know, that immediately makes me worry because that means there's a web page that you visit, and this thing delivers, and we know what a problem people have securing web pages because it just seems that programmers are so sloppy about the code that's used to put up a web page. It's incomprehensible to me that this is a problem today, but it still is. You know, all these web management interfaces are what's constantly being cut through, and here's a security vendor, like a serious security vendor who's got the same problem.

So he says: "For SonicWall NSA devices with SNMP exposed, it's possible to obtain accurate model and version information." You know, SNMP is the network management protocol which exposes an API that allows you basically to access lots of settings in a device. In this case, it's able to obtain model and version information. So he's able to create a correlation. He said: "By comparing the results of my custom method" - his HTML structure reverse engineering - "with the SNMP data from around 5,000 devices," he says, "I've confirmed the accuracy of this detection approach."

So anyway, he then posted a chart showing the lackluster patch status across these devices. The United States has more than half of the globally deployed SonicWall devices. Actually that's a different heatmap. We'll get to that one in a second.

Leo: Oh, sorry. I'm on the wrong heatmap. Well, apologies.

Steve: Yes. But Shadowserver...

Leo: One heatmap looks much like the other.

Steve: Actually, that's a very good point. It is the case. So SonicWall of course, is a U.S. organization. So it's no surprise that the U.S. has more than half of the globally deployed SonicWall devices. There are 390,474 worldwide SonicWall devices. In the U.S., 238,678. So sadly, of the identified global 48,933 currently known vulnerable, still vulnerable since last August, SonicWall devices, 29,107 are detected as still being vulnerable in the U.S. four months after their publisher's and CISA's warning of a 9.8 CVSS vulnerability which is exploitable.

So I say it again, something needs to change. And is it any surprise that ransomware continues to be a scourge across the Internet? On the one hand, any company being victimized with their proprietary data exfiltrated and then held for ransom, you know, that's a crime, doing that to them. That's hacking. But we all know that Internet security can never be a one-and-done install and forget. The connection of an internal corporate network to the global public network is incredibly empowering, but with it comes the

responsibility of managing the security of that interconnection, because that's what you're talking about doing.

You're talking about taking your internal proprietary corporate network, where all kinds of private stuff exists and flows, and interconnecting it to a global network that is jam-packed with bad guys, and they want to get in. So to ever take for granted the nature of the need for security of that interconnection is to risk everything that the organization holds dear. And so I just - it's unconscionable that you could have a SonicWall device like this for which a problem is found in August, and in the U.S. more than 29,000 of them are sitting there just, you know, these two groups, the ransomware groups are just working their way through them.

It feels like the fact that the number is only a hundred, to me that feels like it isn't like a - even though the severity is high, it must be that the exploitability index is low, that is, you know, it takes some work like, you know, pounding at these things in some way in order to get in. But eventually you do. So, boy. Again, to our listeners, just be sure that some sort of email account exists that is being monitored and that is receiving the notifications, you know, that you're on all the equipment vendor notification lists for the equipment that you're using; and that somebody is like, okay, I'll get around to that. No. It's, you know, get that done as a top priority. As I said, something needs to change. I ask why SonicWall isn't just able to go fix this themselves.

Leo: They should be able to push it, shouldn't they.

Steve: Yes. Yes. We have to get there.

Leo: Yeah.

Steve: You know, we're doing it now with consumer routers. It's time to move up to the big iron.

Leo: SonicWall's hardware.

Steve: Yes.

Leo: Okay. Yeah, they should be able to push for updates, yeah.

Steve: It's a top-tier firewall vendor.

Leo: Oh, yeah. Absolutely, yeah.

Steve: Yeah. Okay. So Shadowserver Foundation and Email Encryption, or lack thereof. Speaking of..

Leo: This blows me away.

Steve: Yeah. Speaking of the Shadowserver Foundation, on New Year's Eve morning they posted to their Bluesky Social account. They posted: "We've started notifying owners of hosts running POP3/IMAP services without TLS enabled, meaning usernames and passwords are not encrypted when transmitted. We see around 3.3 million such cases with POP3 and a similar amount with IMAP because most overlap." They said: "It's time to retire those services."

Leo: You've got to wonder if some of them are just being run by individuals; right? No email company would not use TLS.

Steve: Individuals can't. And I'll get to that in a second because all ISPs blocked port 25.

Leo: Right.

Steve: Which is the unencrypted SMTP port.

Leo: Right.

Steve: So can't happen. So this is something we don't talk about often, but it bears reminding everyone. Like the rest of the entire original Internet - meaning web, FTP, DNS, and all the rest - electronic mail exchanged over SMTP, POP, and IMAP protocols was not originally encrypted. It was all sent over simple unencrypted TCP connections in ASCII plaintext, thus making it all completely readable by anyone tapping into any location, whether near to any sender or receiver - such as by an ISP or wireless hotspot operator - or over the public Internet wherever traffic is moving past.

Now, with inertia being the prevailing force that it obviously is on the Internet, we just talked, look at the SonicWall sitting there for four months, patches available, nothing's happening. With inertia being the prevailing force that it obviously is on the Internet, the Shadowserver Foundation reminds us that a sizable portion of email servers have never bothered to move to encryption. You know, no one has ever made them encrypt. Unlike the web with HTTPS where encryption became mandatory, email security has largely fallen through the cracks, even while it has arguably become more important than ever as we depend upon it as our identity authentication of last resort.

That means that all of the email these 3.3 million servers send and receive has remained the same unencrypted plaintext that it was 35 years ago. Right now, today. Those emailed "Oops! I forgot my password" recovery links. The "We just sent you a super-secret 6-digit one-time code to authenticate yourself because it's so important" emails. Those are all out there for anyone to see. And lest we imagine that these 3.3 million email servers must be scattered among backwater countries no one has ever heard of and can't spell, the Shadowserver Foundation thoughtfully provided a heatmap...

Leo: Now? Now you want the heatmap? Now?

Steve: Now we need the heatmap, Leo. Just where these utterly security-negligent machines are located. Guess which country leads the pack?

Leo: Wow.

Steve: Yup. None other than the good old U.S. of A. Within...

Leo: It's not possible that these are misidentified, or they're honeypots, or something like that?

Steve: No, no.

Leo: No? Oh, my god.

Steve: Within our proud borders lie some 898,700 completely unencrypted email servers.

Leo: Unbelievable.

Steve: Those nearly 899,000 email servers are right now, today, this very moment, exchanging email for people who probably have no idea that everything they're sending and receiving is in the clear and readable by anyone who might even be the least bit curious because it takes very little effort. And we know that none of these are people at home, to your point, Leo. We know that they're not at home because long ago ISPs blocked SMTP's port 25 due to rampant spam abuses. So these must be organizations of some size who probably think it's, you know, super spiffy to save some money by running their own email, while apparently never stopping to...

Leo: Thank you for "super spiffy." That's clearly...

Steve: Yeah super spiffy. We've got our own email. You know, we're saving money. That's right.

Leo: Super spiffy.

Steve: Super spiffy. Unfortunately, all the email that they're transacting is readable by anyone. Now, I said there were a total of 3.3 million, and we've accounted for the U.S. taking the top slot at nearly 899,000 instances. So there are others. Germany takes the second spot at 560,900 unencrypted email servers. Poland is in third place at 388,000, followed by Japan at 294,000, and then the Netherlands down to 137,300. Then France, Spain, and you've got to get down to, let's see, France is still over 100,000, Spain at 88,200, and the U.K. at 84.7. So, you know, this is a thing.

Leo: Sheesh.

Steve: Now, having seen these numbers, it would be very interesting to know what is going on. You know, who are these 899,000, Leo, entities in the U.S. who probably run

encrypted web servers with up-to-date TLS certificates because, why? The world insists upon it.

Leo: Ah, yes.

Steve: But they never bothered to think about their email.

Leo: Yup.

Steve: Email servers, just like web servers, connect to each other using the TCP protocol. So just like web servers, it is very possible for email servers to add a layer of authentication and encryption by negotiating TLS certificates with each other. This allows them to each verify the other's identity and to agree upon a shared secret key to use for encrypting and decrypting each other's traffic.

The $64,000 question is how is this ever going to be made to change? Because we know that the phrase "being made to change" is the only way it will ever happen. Web browsers, thanks to the tightly coordinated efforts of the CA/Browser forum, were able to force the entire web server industry to move to encrypted connections by rightfully scaring anyone using a browser that was unable to establish an encrypted connection to a remote web server. At first it was a frightening experience. Today one really needs to work at establishing an unencrypted connection to a web server. You know, I've got to click all sorts of yes, I'm sure, and I know what I'm doing, and my will is updated so, you know, yes, please let me have an unencrypted connection. It's crazy.

So as a consequence, because web browser, you know, nobody wanted to run a server that users would say, uh, I don't think I'm going to go here, and they'd just go somewhere else. Consequently, didn't take long for all web servers to obtain TLS certificates. As we know, this transition to HTTPS Everywhere was tremendously aided by the creation of Let's Encrypt and the ACME protocol, which automated the issuance and installation of free web server domain validation TLS certificates. Unfortunately, nothing like Let's Encrypt exists for email servers.

The ACME protocol is able to verify a server's control over a domain through the presence of a transient signature file located in the .well-known root directory of a web server, or by querying for a TXT record with that domain's DNS. But there is no similar direct support for email servers, despite there being clear demand for it evidenced within Let's Encrypt's feedback forums. People are wanting to encrypt their email. Let's Encrypt says, yeah, we don't do that. Sorry about that.

You know, all of GRC's email transactions are of course encrypted. At the moment, once every year, after I've updated all of GRC's servers with a new certificate from DigiCert, I need to manually reformulate the certificate from binary to ASCII Base64 encoded, and install it into GRC's beloved hMailServer. That's a manual process which I don't mind performing once a year. But as, and if, certificates continue their apparently inexorable reduction in lifetime, any sort of manual process will obviously become increasingly problematic. Since I have multiple Windows and Unix servers that need to be kept synchronized with wildcard domains, this entirely pointless reduction in certificate lifetime will eventually force me to roll my own solution to keep everything running without my intervention.

I've received a great deal of feedback from our listeners who have chimed in with their own issues surrounding shortening certificate lifetimes and the headaches this is creating

for them and for their non-web services because there are many non-web services, and ACME is only used for web services and DNS. Certificates are not used only for the web, you know, and we wish they were being used more for email. But they're used for many other purposes which are being ignored. It appears that the CA/Browser forum is being, I think, somewhat myopic in their apparent belief that the entire world is the web, and thus forcing these short lifetime certificates on everyone.

I've not looked deeply enough into this mess to determine whether it might be possible to delineate the use of short-life certificates only for web services where automation is convenient and supported, while allowing non-web server TLS certificates to remain reasonably multi-year. Alternatively, since we know that web browsers are able to, and have said they would be, eventually independently rejecting any certificate having an out-of-spec total lifetime, meaning the span between "not valid before" and "not valid after" dates, both of which are available.

Browsers have said if that's more than whatever it's supposed to be, like now it's a year, we're just, you know, doesn't matter if it's still valid. If you got it too long ago, we're going to say no. That means that everything could be left as it is, with web browsers being the sole enforcers for short-life web certificates, which would allow everybody else to use longer life certificates.

Anyway, I've wandered well off course here. But my point is, without some means of enforcing the use of TLS certificates for email, history shows us that nothing will ever move these recalcitrant email servers to encryption. If they don't see any problem today, why would they ever make the effort? Especially when it's not particularly easy. And, boy. If we ever get six-day certs, forget about it. The only obvious mechanism for forcing this change would be for those web servers that do support encryption to refuse to accept any insecure email connections.

Leo: Ah. And Gmail could do this with a stroke of a pen because...

Steve: Yes, yes.

Leo: ...they're so big.

Steve: Yes. The problem is, for example, out of fear of missing anyone's important email, I historically configured GRC's email server to accept unencrypted email over port 25...

Leo: It's your fault.

Steve: ...while offering to dynamically upgrade the connection to full security using STARTTLS, which is an SMTP command that allows cooperating email servers to add encryption over a traditionally unencrypted port. But I have to say, now I'm beginning to think that perhaps it's time to end that practice, for GRC to refuse unencrypted email, because another interesting tidbit here is that port 25 has largely become the domain of spammers. Spammers use port 25 because they don't have to have any certs. They can pretend to be anybody they want to be. And there's no verification of their identity which certificates do enforce.

But for those 3.3 million unencrypted email servers in the world, nearly 899,000 of which are in the U.S., before they're going to be able to move to encryption, they're going to need some means of obtaining reasonably priced and reasonably maintained TLS certificates. And that doesn't exist today for small independent servers. You know? It's easy to run an email server unless you have to constantly be updating its certificates. So nobody bothers. It's a mess, Leo.

Leo: I'm shocked because I really thought that every email server now used encryption. I mean, I just - I'm stunned. Do you think these are commercial providers? Or who are these people?

Steve: I really do wonder who they are.

Leo: Yeah. It may well be companies with their own, you know, email?

Steve: Honey, it's those super spiffy [crosstalk].

Leo: Anybody who could have the smarts to configure an email server one would think be able to get a certificate for it. Boy, that's...

Steve: I mean, it is free. If you bring up an email server, and you've got a connection to the Internet, it's free.

Leo: Yeah.

Steve: And I'll bet you that that's how this happened. And because it was working 20 years ago, nobody's revisited it. It's like, well? And they're just not thinking about it. They're, you know, they had to have a certificate for their web server because they probably have a little corporate website; you know? But it isn't easy to do. And we know that, if it isn't easy, and if no one makes them do it...

Leo: No one makes them, that's the key.

Steve: ...they're just not doing it. Yet the employees in that company are receiving password recovery links and...

Leo: Everything. Everything.

Steve: ...6-digit one-time passcodes. Everything. And it's completely in the clear.

Leo: I would love to see yet another heatmap on which servers are being used. Are these primarily Exchange servers? Are they traditional IMAP servers? What are they? You know? SMTP mail? I don't - what are people using? Very wild.

Steve: Okay, a break.

Leo: Break. And more of Steverino coming up in just a bit, including I think the best part of the show I'm waiting for, his...

Steve: I'm saving it for last.

Leo: ...his AI analysis.

Steve: I think I have [crosstalk] to say.

Leo: I'm ready to hear this. He's read all the stuff now. Okay, Steve, on we go with Salt Typhoon.

Steve: So following up on the news, we talked about this last year, which wasn't that long ago.

Leo: Not so long ago.

Steve: This Chinese-backed advanced persistent threat group known as Salt Typhoon had infiltrated all telecom providers. Now three U.S. providers - AT&T, Verizon, and Lumen - all say that they've now evicted Salt Typhoon from their networks. Okay. After this widespread and frighteningly successful hacking campaign came to light, CISA suggested that we should not be relying upon the security of telecom carriers and should instead add our own strong encryption provided by third-party apps such as Signal. Imagine that.

In the aftermath of these attacks, remaining with CISA's recommendation would seem prudent because, you know, who knows whether they actually did evict these guys. And if your traffic happens to cross over some of the telecom carriers that have not yet succeeded in successfully evicting Salt Typhoon, then your communications are still probably not very secure. So if, you know, if you're just ordering pizza, don't bother. But if it's something super sensitive, it's probably worth bringing up something like Signal to hold your conversation.

Also on December 27th the U.S. Department of Health and Human Services issued a Notice of Proposed Rulemaking - god, there's acronyms for everything. We have HHS, Health and Human Services. We also have the Notice of Proposed Rulemaking, that's the NPRM.

Leo: Oh, yeah.

Steve: To modify HIPAA...

Leo: Oh, lord.

Steve: So that's of course HIPAA, the aging Health Insurance Portability and Accountability Act of 1996. So it's been around for a while. Anyway, you could imagine it needs some modernizing. HIPAA regulations will be getting a bunch of new, welcome, and needed cybersecurity rules including the mandatory use of encryption, multifactor authentication, network segmentation - that'll be nice - vulnerability scanning, and more. The show notes went out last night, and I've already seen some of our listeners who had some interesting feedback about this HIPAA change. So I may have some interesting stuff to share from them in follow-up to this next week.

I also got a kick out of this wacky bit. Under the label of "true miscellany," I wanted to mention in passing that the EU, apparently having nothing more pressing to legislate at the moment, which is saying something for the EU, has taken the time to establish USB-C as the official common standard for charging electronic devices throughout their union. There's actually an official document bearing the headline "One common charging solution for all."

In part, the EU legislation reads: "The Commission promotes solutions that favor technological innovation in electronic device charging" - which one would - "while avoiding market fragmentation. The voluntary approach did not meet consumer, European Parliament, or Commission expectations, so we put forward a legislative approach. The common charger will improve consumers' experience, reduce the environmental footprint associated with the production and disposal of unneeded chargers, while maintaining innovation." Wow. In other words, the market didn't settle into any sane and rational standard by itself, so we're going to impose some legislation where needed here.

They said: "The 'common charging' requirements will apply to all handheld mobile phones, tablets, digital cameras, headphones, headsets, portable speakers, handheld videogame consoles, e-readers, earbuds, keyboards, mice, and portable navigation systems as of the 28th of December, 2024, meaning end of last year. These requirements will also apply to laptops as of the 28th of April, 2026."

Leo: Oh, good.

Steve: Yeah. So we have some time with our laptops, even though...

Leo: But I think that's huge. I mean, most of my laptops nowadays use USB charging.

Steve: Exactly.

Leo: But those proprietary chargers just were awful.

Steve: Dumb. "Such transition periods will give industry sufficient time to adapt" - which would be nice - "before the entry into application. The main elements are as follows: A harmonized charging port for electronic devices. USB-C will be the common port. This will allow consumers to charge their devices with any USB-C charger, regardless of the device brand. Harmonized fast-charging technology: Harmonization will help prevent different producers from unjustifiably limiting charging speed and will help to ensure that charging speed is the same when using any compatible charger for a device.

"Unbundling the sale of a charger from the sale of the electronic device: Consumers will be able to purchase a new electronic device without a new charger. This will limit the number of chargers on the market or left unused. Reducing production and disposal of new chargers is estimated to reduce the amount of electronic waste by 980 tons yearly." Wow.

Leo: Wow.

Steve: 980 tons' worth of chargers eliminated. No more drawers full of unneeded, unwanted, unused, and forgotten chargers. So before long those in the EU will be spared the experience of opening the box and thinking: "Oh, shoot, not another damn charger."

They did note that since the wireless magnetic induction charging market is so far behaving itself and is not showing undue fragmentation, they did not feel the need to impose any order there. But that market, too, might need some harmonization if things start going all wild and woolly. So they're keeping a watchful eye on it. They just wanted everyone to know, now, you guys, behave yourselves over there in the magnetic induction side.

And we have the DOOM CAPTCHA. That's right. Since nobody likes CAPTCHAs, an enterprising software engineer has created a DOOM CAPTCHA system where you have to kill at least three bad guys in the DOOM video game to proceed to a website. And it's actually a functioning CAPTCHA. Since I thought our listeners would get a kick out of it, I gave it one of GRC's shortcuts of just "doom." So grc.sc/doom will take you to a doom-captcha.vercel.app. And its author wrote: "A CAPTCHA that lets you play DOOM to prove you're human," and he said, "for educational and entertainment purposes."

He said: "The project works by leveraging Emscripten to compile a minimal port of Doom to WebAssem and enable intercommunication between the C-based game run loop, which is g_game.c, and the JavaScript-based CAPTCHA UI. Some extensions were made to the game to introduce relevant events needed for its usage in the context of a CAPTCHA. Started out with a minimal SDL port based of Doom that can be efficiently compiled to WebAssem, then tweaked the build to make it compatible with the shareware version of wad - that's doom1.wad - for legal use."

Leo: You know, any computer can kill three monsters in Doom. That is the worst CAPTCHA ever.

Steve: Actually, yes. I'm no videogamer, Leo. So I was promptly killed right off the bat while I was working out the arrow keys and the spacebar.

Leo: Oh, right.

Steve: For movement and firing.

Leo: You're just better at it than a human.

Steve: It's not that difficult to kill three baddies because I was - even I was able to pull that off on my second try. Anyway, since, as I said, grc.sc/doom. One of the people who

received the show notes last night sent me a note and said, "I thought I remembered this from the past, and I think it was maybe Episode 8 - it was 890 something," he said, "where we talked about this." I don't know whether this is exactly the same or whether this has been updated to be using WebAssem. But, you know, I mean, it does run in a browser. And one of these, you know, boy, if I got into WebAssembly, I would be dangerous, I think, because, you know, mix my assembly language interest...

Leo: This isn't that easy, is it.

Steve: It's not that easy. Now, what I did was I just stood there, so they come out right there.

Leo: Yeah, you shouldn't go to them. That's right.

Steve: Yes, exactly.

Leo: Yeah. There's one.

Steve: I meant to kill the three just by...

Leo: Oh, he's got me. Oh.

Steve: Yeah.

Leo: Oh, this is harder than it looks. There we go. There we go. Oh, ho.

Steve: [Crosstalk] solve it. Yup. Look what I got.

Leo: That is not good. Any computer will play this better than you will, I promise.

Steve: Yeah.

Leo: That's hysterical.

Steve: I think that's true.

Leo: Yeah.

Steve: Okay. So we're ready to go to AI Training and Inference. We have one last break.

Leo: Yes.

Steve: So let's take that, and then we'll plow in.

Leo: All right, Steve. I am dying to hear...

Steve: Okay.

Leo: What you think about all this AI stuff.

Steve: So as I said at the top of the podcast, and I will reiterate, Security Now! will not be evolving into "AI Today."

Leo: No. We have shows for that. That's fine.

Steve: Yes. And that said, aside from the fact that the recent truly astonishing advances in AI are going to directly impact everyone's lives outside of the security sphere, I'm also very certain that we're going to be seeing AI's impact upon the security of our software and operating systems, and we may not be needing to wait long. So over the course of the next few years, I'm sure that the topic of AI will be reemerging. And I'm not saying I'm never going to talk about it again because, you know, it'll just be fun to talk about the major advances that I expect that we're going to be seeing, one actually I'll be talking about in a second, only about a month away.

So our listeners have been following my journey through this topic, and it's not been a straight line. More than anything else, I endeavor to be an honest researcher. An honest researcher will readily revise their entire belief system as required when presented with new facts and information. Clutching to obsolete dogma simply because it's familiar and comfortable is not the way of science. And it was because I was puzzled and confused by what I was experiencing firsthand that I went searching for that information. I believe I've found it. I believe I understand it, at least as much as is possible without actually implementing it myself; and I've got other work to do, so that's not going to happen. And I've been changed by what I learned.

Three weeks ago, as I said, I might have something to say about this before we met again today. And I said, if so, I would probably enjoy sharing that with this audience with a special email over the holidays. Now, the possibility of that happening induced more than 1,100 of our listeners, who had not already signed up to the Security Now! mailing, to do so. So for that reason alone, due to the declaration of interest, I felt I had to say something. Today, I have much more to say on the topic than I did nine days ago, last Monday, December 30th, when I sent that out. But let's start with what those 15,060 subscribers received from me last week, then I'll expand a bit on what I think are the most important points and what I've continued to learn since.

So what I wrote then was: "When I first set about writing this email, my plan was to share what I had learned during the first half of our three-week hiatus from the podcast. But it quickly grew long, even longer than this, because I've learned quite a lot about what's going on with AI. Since I suspect no one wants to read a podcast-length piece of email which I would largely need to repeat for the podcast anyway" - which is what I'm doing now - "I'm going to distill this into an historical narrative to summarize a few key

points and milestones. Then I'm going to point everyone to a 22-minute YouTube video that should serve to raise everyone's eyebrows."

So here it is. First, everything that's going on is about neural networks. This has become so obvious to those in the business that they no longer talk about it. It would be like making a point of saying that today's computers run on electricity. Duh.

Okay. AI computation can be divided into "pre-training" and "test-time," also called "inference-time." Pre-training is the monumental task, and it is monumental, of putting information into a massive and initially untrained neural network. Information is "put into" the network by comparing the network's output against the expected or correct output, then back- propagating tweaks to the neural network's vast quantity of parameters to move the network's latest output more toward the correct output. A modern neural network like GPT-3, which is already obsolete, had 175 billion parameters interlinking its neurons, each of which requires tweaking. This is done over and over and over, many millions of times, across a massive body of "knowledge," which I have in quotes, to gradually train the network to generate the proper output for any input.

Counterintuitive though it may be, the result of this training is a neural network that actually contains the knowledge that was used to train it. It is a true knowledge representation. Now, if that's difficult to swallow, consider human DNA as an analogy. DNA contains all of the knowledge that's required to build a person. The fact that DNA is not itself intelligent or sentient doesn't mean that it's not jam-packed with knowledge. In fact, the advances that have most recently been made, which I'll get to in a bit, are dramatic improvements in the technology for extracting that stored knowledge from the network. That's why I titled today's podcast "AI Training and Inference." The inference is the second half.

The implementation of neural networks is surprisingly simple, requiring only a lot of standard multiplication and addition, pipelined with massive parallelism. This is exactly what GPUs were designed to do. They were originally designed to perform the many simple 3D calculations needed for modern gaming. Then they were employed to solve hash problems to mine cryptocurrency. But now they lie at the heart of all neural network AI.

Now, even when powered by massive arrays of the fastest GPUs rented from cloud providers, this "pre-training" approach has become prohibitively, well, was becoming, and is, prohibitively expensive and time consuming. But seven years ago, in 2017, a team of eight Google AI researchers published a truly ground-breaking paper titled "Attention is all you need." The title was inspired by the famous Beatles song "Love Is All You Need," and the paper introduced the technology they named "Transformers." Actually, it was named that because one of the researchers like the sound of the word.

The best way to think of "Transformer" technology is that it allows massive neural networks to be trained much more efficiently in parallel. This insightful paper also introduced the idea that not all of the training tokens that were being fed into the network, which is the long string of data being fed into a model during one training iteration, not all of those tokens needed to be considered with equal strength because they were not all equally important. In other words, more attention could be given to some than others. These breakthroughs resulted in a massive overall improvement in training speed which, in turn, allowed vastly larger networks to be created and trained in reasonable time.

Basically that paper allowed - it solved the problem that they were hitting five years ago, six and seven years ago, that it just - training took too long. That limited the size of the networks, so that limited the quality of the networks. What happened was it then, thanks

to this breakthrough, it became practical and possible to train much larger neural networks, which is what gave birth to today's LLMs (Large Language Models).

Now, the GPT in ChatGPT stands for Generative Pre-trained Transformer. Pre-trained is the training; transformer is this technology. But over time, once again, researchers began running into new limitations. They wanted even bigger networks because bigger networks provided more accurate results. But the bigger the network, the slower and more time consuming, and thus costly, was its training. It would have been theoretically possible to keep pushing that upward, but a better solution was discovered: post-training computation.

Traditional training of massive LLMs was very expensive. The breakthrough Transformer tech that made LLM-scale neural networks feasible for the first time, well, now that was being taken for granted. But at least the training was a one-time investment. After that, a query of the network could be made almost instantly and, therefore, for almost no money. But the trouble was that even with the largest practical networks, the results could be unreliable, known as "hallucinations." Aside from just being annoying, any neural network that was going to hallucinate and just make stuff up could never be relied upon to build chains of inference where its outputs could be used as new inputs to explore consequences when seeking solutions to problems. Being able to reliably feed back a network's output into its inputs would begin to look a lot like thinking, and thus inference for true problem solving.

Then, a few years ago, researchers began to better appreciate what could be done if a neural network's answer was not needed instantly. They began exploring what could be accomplished post-training if, when making a query, some time and computation, and thus money, could be spent working with the pre-trained network. This is known as "test-time computation," and it's the key to the next level breakthrough.

By making a great many queries of the pre-trained network and comparing multiple results, researchers discovered that the overall reliability could be improved so much that it would become possible to create reliable inference chains for true problem solving. Using the jargon of the industry, this is often called "chains of thought," although I still object to, you know, giving too much credit, imbuing these with too much human brain technology.

Leo: Yes, yeah. Thinking involved.

Steve: So inference chains would allow for problem-solving behavior by extracting the stored knowledge that had been trained into these networks, and the pre-trained model could also be used for the correction of its own errors. Now, I should note that the reason asking the same question multiple times results in multiple different answers is that researchers also had long ago discovered with neural networks that introducing just a bit of random noise, which is called "the temperature," into neural networks resulted in superior performance. And yes, if this all sounds suspiciously like voodoo, you're not wrong, but it works anyway.

OpenAI's recently released o1 model, which I talked about at the very end of last year, is the first of these more expensive test-time inference-chain AIs to be made widely available. It offers a truly astonishing improvement over the previous ChatGPT 4o models that we were using. Since o1 is expensive for OpenAI to offer on a per-query basis, subscribers are limited to seven full queries per day. But the o1 mini model, which is faster and still much better, but not as good, can be used without limit.

But wait. There's more. The big news is that during their celebration of the holidays, OpenAI revealed that they have an o3 model that blows away their brand new o1 model. It's not yet available, but it's coming soon. What IS available are the results of its benchmarks, and that's why I believe you need to make time to watch this YouTube video. I created a GRC shortcut with this episode number, which is 1007, so grc.sc/1007. That will bounce you to a, I think it's 22-minute YouTube video talking about the benchmarks that have been the independent benchmarks that have been run against this o3 model.

Okay. So is it AGI? OpenAI is saying "not quite," but there's little question that they're closing in on it. As you'll see in that video, the performance of OpenAI's latest o3 model, when pitted against independent evaluation benchmarks designed specifically to measure the general reasoning strength of AIs - when confronted by problems that were absolutely never part of the AI's training set - demonstrate reasoning abilities superior to most humans. You need to watch the video: grc.sc/1007.

Even if it were AGI, even if it were AGI, and we'll probably get not far from that, people are saying it is, I don't care. But that doesn't mean it's taking over. The "AGI" designation is only meant to indicate that over a wide range of cognitive problem-solving tasks an AI can outperform a knowledgeable person. Computers can already beat the best chess, Go, and poker players. I think it's very clear that today's AIs are not far from being superior to humans at general problem solving. That doesn't make them Frankenstein's monster to be feared; it only makes AI a new and exceedingly useful tool.

Many years ago I grabbed the domain "clevermonkies.com" just because I thought it was fun. It occurs to me that it takes very clever monkeys indeed to create something even more clever than themselves. All the evidence I've seen indicates that we're on the cusp of doing just that.

Okay. So that, with a little bit of editing to improve it, that's what our listeners received from me over the holidays. If you take nothing else away from this discussion of AI today, here is the one point I want to firmly plant into everyone's mind because this is the sticking point that I see everywhere. Nothing that was true about this field of research yesterday will remain true tomorrow. Nothing. This entire field of AI research is the fastest moving target I have ever experienced in my nearly 70 years of life.

There are a number of consequences to this fact. For one, no book about AI that was written a year ago or six months ago, or even last month, will be usefully up to date about what's happening today. Books written in the past can definitely be useful for describing the history of AI, and as a snapshot of a point in time. But even their predictions will prove to have been wildly wrong. The guys at OpenAI who are working on this and ought to know, believed two years ago that at least another decade, another 10 years, would be needed to achieve what they announced last month and are getting ready to unveil. They thought it would take 10 years. It took two.

One of the factors in facilitating this astonishing speed of development is that it turned out that much of what was needed was scale, and a weird side effect of cloud-side computing is that it's massively scalable. If you can pay to rent it, you get to use it. So investor dollars were pumped into the training of ever more complex models, and they kept seeing surprising improvements in performance.

Leo's original appraisal of Large Language Models as fancy spelling correctors was an accurate and useful from-the-hip summary of OpenAI's ChatGPT-3 model. That's their take on it, too. ChatGPT-3 produced grammatically correct language, but it only coincidentally and occasionally produced anything highly meaningful. If it was left to keep talking, it would soon get lost and wander off course to produce grammatically correct nonsense.

Even so, back then, highly creative people who operate on the cutting edge, like MacBreak Weekly's Alex Lindsay, were using the ChatGPT-3 model as a source of new ideas and inspiration. As I wrote this I was reminded of how popular formal brainstorming once was, where sometimes random ideas were just tossed out without any filtering, and that was the entire point, to say something as a means of inspiring some new perspective. So even ChatGPT-3 was useful for the nonsense that it sometimes produced.

But as a consequence of everything I've learned over the past three weeks, and of the events which have transpired since, our previous podcast title, Podcast 1005, three weeks ago, "The Wizard of Oz..."

Leo: How quickly that ages, huh?

Steve: ...no longer seems, yes, no longer seems to fit, and I'm a bit embarrassed by what I wrote because it no longer reflects reality. As I said earlier, an honest researcher may need to discard previous belief systems when confronted with new information and facts. Never has that been more true than it is here. I'm needing to continuously update my own internal model.

There is an unfortunate downside emerging, however. Unfortunate, I suppose, but inevitable. With startling speed, AI has moved from a curio in the corner of university and corporate R&D labs into big business. That meant that the suits in their neckties with their non-disclosure agreements descended upon the labs of the once freely and fruitfully collaborating academia-oriented researchers and dropped the cone of silence over their ongoing work.

In the Distinguished Lecture Series at the Paul Allen School, one of OpenAI's leading researchers, Noam Brown, gave a lecture titled "Parables on the Power of Planning in AI: From Poker to Diplomacy." I have a YouTube link to Noam's excellent talk at the end of the show notes. During his lecture you could so clearly see Noam's unbridled enthusiasm and love of his subject, and also his disappointment when he was forced to stop himself short to prevent sharing some detail of his work that was now deemed to be proprietary and no longer his to share.

We only have Google's breakthrough Transformer and Attention technology - which was the sole enabler of the subsequent LLM revolution - because seven years ago, back in 2017 when things were still moving somewhat slowly, Google AI researchers were freely publishing their work as the academic curiosity that it was at the time. They were working on improving Google's inter-language translation capabilities, and this inspiration emerged unbidden from a chance meeting of eight Googlers from various parts of the organization. Would such a breakthrough be published in today's climate? Seems unlikely.

And now OpenAI is seeming less open than it once was. We know that ChatGPT-3 used a neural network containing an astonishing 175 billion neuron-interlinking parameters, the 10 digits of accuracy each. We know that because OpenAI freely told us. But we have no similar information about any of their succeeding models. The sizes of the various ChatGPT-4 models, not to mention o1 and o3, have become closely held secrets - as have details of their operation.

Leo: This is something that Elon's been complaining about; right? This is why he's suing them.

Steve: Yup.

Leo: Yeah.

Steve: He said: "Fortunately, a massive amount of detail - all detail needed for recreating much of what we see today from the corporate side - had previously been shared in the public domain, and research continues with new vigor and doubtless with new funding within academia. And remember that it wasn't so long ago that Apple was getting patents on Andy Hertzfeld's clever stepwise circle drawing algorithms for bitmaps. Very little of anything that's really useful remains secret forever, and it seems clear that before long we're going to have AI everywhere."

Okay, now, I would love to spend more time talking about the way neural networks function in detail because there are some very cool aspects of that, too. But that's not the purpose of this podcast, and perhaps I'll find another opportunity for that in the future. There are absolutely already tons of videos on YouTube talking about all of this for anyone who's interested, and YouTube's recommendation engine appears to be quite excellent. Because as soon as I started digging around in there, I got a lot of great points.

Leo: There's a lot of good stuff, yeah.

Steve: Yeah. I do need to point out a specific series of astonishingly well-conceived and produced instructional videos on this topic from a guy named Grant Sanderson.

Leo: Oh, I've watched these. They are really good.

Steve: Oh. Oh.

Leo: This was how I got my education in this stuff, yes, I agree.

Steve: Grant's website is 3blue1brown.com, and Grant's bio says: "These videos, and the animation engine behind them, began as side projects as I was wrapping up my time studying math and computer science at Stanford. After graduating, I worked for Khan Academy producing videos, articles, and exercises, primarily focused on multivariate calculus. Since the end of 2016, my primary focus has been on 3blue1brown and its associated projects. In those years, I've also had the pleasure of contributing to a number of different outlets for math exposition, including spending a semester lecturing for an MIT course on computational thinking, contributing a Netflix documentary about infinity, writing for Quanta, and collaborating with many other educational YouTube channels." I have to say his animated visualizations...

Leo: They're very good, yeah.

Steve: ...are astonishing.

Leo: This is the one I found the most useful, if you just want a quick introduction. He put it out in November, "LLMs for Beginners." Very good, very - really well done. And knowledgeable.

Steve: Yes. I have a link in the show notes. He did a series of eight which starts on neural networks and runs through all of this technology - transformers, back propagation, the whole breakthrough of attention and how that operates. Anyway, I recommend them without reservation to anyone who's interested in understanding more of the inner workings of the comparatively, and I love the word, "ancient" technology of neural networks because this stuff's been around forever.

Now, what's interesting about this is that this old technology of neural networks has recently been given new life thanks solely to the scalability of cloud-based computing and the presence of GPUs which are able to perform massive amounts of simple computation operations. So long as we have sufficient power, it appears - now, processing power, and as we know, electrical power, too - it appears that the world is facing, I believe, a true breakthrough, thanks to the scale of compute and training we've been able to throw at the problem.

However, what we have today works and is working, but it is incredibly inefficient. It works only due to the massive scale we've managed to throw at neural network technology, which is itself an extremely flexible but inefficient technology. For example, it's possible to train a neural network that has just a handful of neurons to perform a simple binary adder function. But the same thing can be done far more efficiently with a couple of logical NAND gates. The thing that makes the handful of neurons potentially more interesting is that the same network could be trained to perform other simple functions. But the fundamental problem remains that any simple function that a neural network could be trained to do could be reduced to a far more efficient couple of NAND gates.

So here's what I think will eventually emerge someday. And I have no idea whatsoever when that might be. My hunch is that, just as with the handful of neurons that can be trained to perform simple logic functions, we're going to eventually discover that there is a far simpler way to solve the same AI implementation problems much more efficiently than we're currently solving them by throwing massive scale of inefficient neural networks at the problem. I have no idea what that solution might be.

But the intriguing thing here is that cognitive science researchers now have a crude sort of brain that does manage to store a useful amount of knowledge and is able to use that knowledge to solve novel problems and, I suspect before long, to truly invent new things. People are already beginning to ask, looking at these networks, exactly how it does this because, believe it or not, that remains a mystery. What is no mystery is what transpires here every Tuesday as it will next Tuesday and for many more Tuesdays to come.

Leo: You know, I like your idea that it might be not simply throwing more power at the existing structures, but finding a new structure that might be more efficient. There is a - I sent you a link. There is an article that came out five years ago by this guy, who is a well-known researcher in reinforcement learning and AI. And he actually had an insight. It's kind of funny. He had an insight back in 2019, he calls it the Bitter Lesson. He says: "The biggest lesson that can be read from 70 years of AI research is that the best way to make AI better is to give it more power." Because of Moore's Law, that's what we're seeing.

Steve: Yup.

Leo: It's more power. So he says the other, the second general lesson is the actual contents of minds are - our own minds, right - are tremendously, irredeemably complex. So let's stop trying to find simple ways to think about the contents of minds. That's probably the wrong thing to try to do, to duplicate the human mind. We want AI agents that can discover like we can, can learn like we can so that we don't have to reproduce the complexity of our own minds. We can let them learn.

Steve: Yeah, that's really what happened is, you know, neural networks are interesting because they're self-organizing. And when, like when you train a multilevel neural network that has, like, three or four layers of interconnected neurons to do image recognition, it turns out you're able to do it. It's able pretty easily to recognize handwriting, and that works when you give it a whole bunch of samples. But then you look at how it's doing it, like what do the individual layers of neurons hold.

Leo: We have no idea.

Steve: And it's just it looks like noise.

Leo: Yes.

Steve: It's just junk.

Leo: Yes.

Steve: And it's like, you know, how is it doing this, and we don't know. And believe me, Leo, when you're talking about even ChatGPT-3, that is now a comparatively simple old technology from oh, gee, 90 days ago, and 175 billion neurons?

Leo: Yeah.

Steve: We have no idea. You know, it comes out, and we, it's like, whoa, look at that, it works. We don't know why.

Leo: We don't know what's going on in there.

Steve: No.

Leo: It's a black box. I'm very excited. I do think that, I mean, you know, look, Sam Altman's a great marketer and a great showman. But I do think that he has something that we're going to see in the next few months, that is probably as close to AGI as we need to get.

Steve: Yes. Yes. I think that's absolutely right. I'm worried about what it's going to cost because I probably want to use it, and it looks like it's going to be expensive. You know, there's like a Pro version of o1.

Leo: Two hundred bucks. He says they're losing money on the Pro version at 200 bucks a month because people are using it so much.

Steve: Yeah. But let's hope they can make it up in quantity.

Leo: I have a friend who works in the business who took me aside some months ago and said, "The next decade is going to look very weird." It just is what you said. It's moving so - it's faster than anything we've ever seen.

Steve: Yes. Yes.

Leo: And the developments that are going to happen over the next few years even are mind-bending.

Steve: Yes. I would advise anyone listening when anyone asks them what they think about AI, they can say, well, I'll tell you what I thought last month.

Leo: Yeah.

Steve: Because, I'm not kidding you, it is a shockingly fast-moving target. And the reason is it turns out there was an infrastructure ready to scale.

Leo: Yes.

Steve: There was infrastructure...

Leo: That's the key.

Steve: ...waiting for AI.

Leo: And then, yes, and Moore's Law has scaled it so fast. So just so you feel reassured you do not have to become the AI Show, at this point I'm probably going to rechristen This Week in Google to This Week in Intelligent Machines because I think that's really the most interesting development for this year and the years to come. And Google has become less and less interesting as a single company. But what's happening in all of those companies is more interesting.

Steve: Well, that's good because that's also This Week in IM.

Leo: Yeah. I like it; right? TWiM. Intelligent Machines I thought was better than AI.

Steve: So tell me about Elon because I'm not up to speed on his...

Leo: It's hard to know what his reasoning is. But he has sued now OpenAI because he says, you know, our original concept, it's true, the charter, founding - he was a founding member.

Steve: Was it to be open.

Leo: Was it to be open. And he said in the beginning no company should control artificial intelligence. And so he's suing them because they want to eliminate their nonprofit status, and they're converting to a fully for-profit. Although it might be a public benefit corporation. Nevertheless, Elon's right on the surface that it shouldn't be controlled by any big company. You might say if you were cynical that he's really just trying to slow OpenAI down so his own corporate commercial for-profit AI, Grok, can catch up. I think that might be closer to the truth. You never know with Elon. But I think on the surface he's right. No big company should support, should be in control of this. This needs to be something we all use. And it saddens me when I hear a scientist, because of an NDA, say, "Oh, I can't tell you what I'm doing."

Steve: Yeah. You probably heard that there was a paper out of China also where they believe they've figured out how o3 works, even though OpenAI is not saying.

Leo: Interesting. Yeah. That's the good news is that this is such a game change that I think every country, every scientist, everybody's working on this. And it's going to be a very interesting time we're in. I don't know if it's going to be a good time. But it's going to be interesting.

Steve: Yeah. Well...

Leo: It's [crosstalk] disruptive.

Steve: Well, as I said, I got into this because I started using it as sort of a super Internet search engine, and...

Leo: Right. It's good for that.

Steve: It is very useful.

Leo: Very good for that.

Steve: It is very useful. You absolutely have to check its work because it does, you know...

Leo: The best ones give you references that you can follow back.

Steve: Yeah.

Leo: I use Perplexity AI for my search research. And it's always very good about, first of all, it's very up to date, unlike some of the older models. Its training continues.

Steve: Well, and I did ask, I think it was 4o, because I asked it something that it didn't seem right. And I said, "When did your training stop?" And it said, "I stopped in October of 2023."

Leo: Yeah, yeah, said a date, yeah.

Steve: Okay, well, then, you don't know what I'm asking you.

Leo: Exactly. Exactly. So OpenAI does have a GPT that is connected to the Internet. But Perplexity's I think is the best. It's not only a very good model, but it's...

Steve: I'm hearing that Claude is also very good.

Leo: Claude's very good, too.

Steve: For proposed stuff.

Leo: Claude has, yeah, Claude has a search tool. I do think this is going to replace search. I have stopped using traditional search entirely.

Steve: Yeah. And you have to know that's where Google is putting so much of their effort.

Leo: They seem a little behind. Anyway, it's going to be a very, very interesting time, shall we say. And you don't - while I want you to continue to cover AI to whatever extent you wish, just be reassured AI is absolutely the focus of a number of our shows, and especially I think This Week in Google's going to become more of an - it already is a lot about AI.

Steve: And no one better than Jeff to steer the ship.

Leo: Well, I'll put my two cents in, too. And one of the things we're going to do as we transform that show is to bring in experts because we need expert information.

Steve: Neat.

Leo: Yeah, I think that's going to be very fun.