



SuperDataScience

SDS PODCAST
EPISODE 1040:
CYBERSECURITY FOR
THE AGENTIC AI ERA,
WITH AMY
BLACKSHAW



- Jon Krohn: 00:00 For four and a half days, autonomous AI agents moved through Hugging Faces infrastructure at one action every 20 seconds. How did they get so far? A plain text password and a single master key that opened 136 doors. Welcome to another episode of the SuperData Science Podcast. I'm your host, Jon Krohn. Today's guest is Amy Blackshaw, a senior vice president at Palo Alto Networks, the world's largest independent cybersecurity company. In this episode, Amy explains why Frontier AI has made the thinking part of a cyber attack nearly free, thereby making cyber attacks inexpensive, fast, and relentless like we saw with the recent Hugging Face breach. Amy magnificently explains cybersecurity and how to protect ourselves and our businesses from cyber threats in the agentic era that's here and that's coming. It's a riveting episode. Enjoy. Amy, welcome to the Super Data Science Podcast. How's it going today?
- Amy Blackshaw: 01:01 It's good, Jon. Thanks so much for having me. I'm excited to be here.
- Jon Krohn: 01:04 Yes, yes. We're excited to have you as well. You work for a company that is actually gigantic. I was just looking at its stock price before we started recording over \$300 billion in market valuation. It's called Palo Alto Networks. And my understanding is that for anybody who works in cybersecurity, they will already know what Palo Alto Networks is, but not all of our listeners work in cybersecurity. So do you want to tell us about the company?
- Amy Blackshaw: 01:32 Yeah, absolutely. So yes, we are a very large cybersecurity company, the largest independent cybersecurity vendor in the world, and we're the global AI security leader. So we protect over 75,000 customers in the way that they go about in their business. And we offer a portfolio of cybersecurity solutions spanning across network security, cloud security, security operations, AI security, and identity security. We also have a unit 42, which is our



threat intelligence, which helps customers eliminate complexity and help empower our customers to modernize and ultimately secure themselves so they can move fast in what they need to

Jon Krohn: 02:20

Do. Nice. Let's double click on that identity security that you mentioned there. What is that?

Amy Blackshaw: 02:26

Yeah. So the simplest way to think about identity security is that every breach today that you read about or that you see in the news or that you've heard from a peer or a friend comes down to the idea that an attacker has gotten hold of an identity that they should not have had access to. So whether that's a login or a key or a token, and then they use that to move. So they don't have to break encryption, they don't have to do some zero day attack that you might hear in movies or different wild ideas of what attacks look like. It's just becoming someone or something that the system has already trusted. And once that happens, everything downwind of that identity is at risk. So identity security is this sort of discipline, this idea that you have to decide who and what is allowed to act inside of a company and how long they have access to do what they need to do.

03:29

And for years it was treated as plumbing. We are now moving in a different motion with AI and the threat landscape that we live in, and it's all about securing every identity with the right level of privilege control.

Jon Krohn: 03:40

And it sounds like you have a specific solution to this identity security problem. It's called IDERA, which seems to be related to an acquisition that Palo Alto Networks made, a company called CyberArk.

Amy Blackshaw: 03:52

That's right. CyberArk was the pioneer and the founder of the privileged access management space 26 years ago and over time became the identity security leader. Palo Alto Networks acquired CyberArk as of February in 2026. So I



was part of the CyberArk team that was joining Palo Alto Networks and we have rebranded the CyberArk platform into IDERA. So the IDERA platform secures every identity, whether it's a human, a machine, an AI agent with that right level of privilege controls. And we're very excited to be part of the Palo Alto Networks organization because we are being infused with resources, innovation, scale, speed to bring more technology to our customers to help them secure their identities.

Jon Krohn: 04:43

Nice. Thanks for giving us the idea behind Idira. And if people are looking to know how to spell that, it's I-D-I-R-A, IDIRA. So let's talk about why identity security is now so much more important in this agentic AI era. So what has changed about how attacks unfold now that frontier AI models can chain security vulnerabilities together all on their own?

Amy Blackshaw: 05:09

Yeah, I think the part that surprises a lot of people is that attacks themselves didn't change. We didn't wake up with Frontier AI and just say, "Wow, we now have all of these new vectors that we've never seen before." It's that the techniques that an AI-driven attacker are the same ones that we have known for years and for decades in some cases. But what has changed is the economics. So for the entire history of the cybersecurity industry, the expensive part of the attack was sort of the thinking, right? How does one get in? How does a skilled human have to understand their target environment, work out how to look at one vulnerability in one system, and like you said, sort of chain them to understand how a foothold could be turned into something catastrophic for the target organization? Sometimes that took weeks, it took experts.

06:06

There was an expense in that attack. Now, Frontier AI has deleted that cost. The thinking is now it's instant, it's tireless, it's exhaustive when you're talking about these agentic models. It can check every door, not just the open



obvious ones, and it's happy to do that for a small company or a large company.

06:31 I think I want to take a quick second to be very specific about something. It's about who the villain is because the villain isn't the models that are enabling this tireless approach of attack vector chaining. The villain is still the attacker. They just now have the same capabilities that we do and they're operating under a new economics. So the entry is cheap and enumeration is free. Yeah,

Jon Krohn: 06:56 There's all kinds of wonderful positive things about abundant, very inexpensive intelligence. You can imagine an education in food security, in healthcare, there's all kinds of positive things, but that same intelligence can be used by bad actors.

Amy Blackshaw: 07:15 Exactly right.

Jon Krohn: 07:16 So speaking of which, something that we did an episode focused on this, but you are a security expert. And so I'd love to hear from your perspective, what happened in July with OpenAI's agents getting into the hugging face systems? What did the agent do and how did it reach cluster admin authority in a matter of hours?

Amy Blackshaw: 07:41 Yeah. It's an incident that I think has gotten well deserved coverage because I think it really helps provide a clear example of frontier AI models and sort of this idea of machine speed and how it can play out in an actual incident. So they were dealing with an intrusion into their production infrastructure, hugging face, where the attacker was inside for four and a half days and it carried out 17 and a half thousand attacker actions. That's about an action every 20 seconds sustained. A human can't do that. That is machine speed and that is what persistent machine speed looks like. So these autonomous AI agents, they escaped the sandbox that they were being tested in. They reached the internet, they commandeered



a third party server, and then they got to work. But this is what the important point is. Speed got them through the door, but speed didn't get them to the critical resources because to achieve their goal, an attacker, whether it's the swarm of AI agents or just a human, they need to get access to an identity and the resources and permissions that are associated with that identity.

08:56 So an attacker doesn't care if it's an identity that belongs to me, Amy Blackshaw, or if it's a service account or an AI agent. If they can use that identity to get where they're going, it is valuable to them. So that's what we saw here. Those four and a half days of continuous exposure wasn't just some incredibly sophisticated AI technique. It was actually fundamentally a problem of identity and privilege hygiene that was non-existence in this infrastructure. So there was this plain text password sitting where the process could access it. There was one master key that opened 136 doors and there was an unreviewed service account that had standing admin privileges across two clusters. So what does that mean to people who maybe aren't in cybersecurity? It's like you lock the front door, but you make it really easy to get in to somebody who's going to be persistent with a pick lock.

Jon Krohn: 09:56 The old key under the front mat.

Amy Blackshaw: 09:59 Exactly. But this one's even worse. They open the door with the key and then on the kitchen table are all of the safe keys. They just laid them out and said, now that you're in here, you might as well get into where all of our crown jewels are or where all of our silverware and things that are jewelry that are really important to you. So none of the weaknesses here required a sophisticated AI exploit. None of them did. It was getting in that was enabled by AI models for sure and that machine speed. But what essentially happened was that privilege was left standing and that it was never reviewed and that was taken advantage of by the attacker. It's the tail as old as



time. They get access, they elevate privileges, they move laterally, they elevate privileges, and then they get to where they're going.

10:53 It is a hygiene of identity security that could really change the outcome in an attack like this.

Jon Krohn: 11:00 Excellent. We'll talk more about privilege and hygiene. And I've got to say, anytime we have some big AI cybersecurity incident, we're going to have to get you back on the show because that was by far the best explanation. I've heard of that incident having read about it dozens of times. Amy, that was the best one for sure. So thank you for coming and sharing that with us. Before we get back to privilege, something that came up in our research for this episode that I don't even really understand why this is so important, so I'm looking forward to being illuminated about this, but you've cited a 72 minute breakout time in real cybersecurity incidents versus 25 minutes in simulations. So less than half the time in simulations relative to real incidents. Where do those numbers come from? What is the significance of those numbers and what can a defender do in that window?

Amy Blackshaw: 11:55 Yeah, it's a great question. So to think about it, first they come from our unit 42 specialists who every day are holding hands with our customers and helping respond to real incidents. And so what they see is that we measure an attacker going from sort of first entry point or foothold into real impact, whether that's breach or wherever that attacker's trying to go in about 72 minutes. In simulations where we are able to do testing to simulate these type of incidents, it's closer to 25. And so when we're talking about that speed, even 72 minutes, there's a few things that come to mind. So first and foremost, the idea that a human-led incident response is going to be able to identify an attack, see the depth of it, and then respond even within that 72-minute window is hard. 25



minutes is really hard. And so we believe that security teams need to be fighting AI with AI and by using AI within their own security teams to not be on their back foot when the attacker's leveraging AI, they need to be leveraging AI.

13:08 And so using the different frontier models within their security tools, which Palo Alto Networks does, and enabling them to respond with agentic workflows so that you obviously have humans in the loop in security teams, but that you are leveraging AI to fight AI. Number two, virtual patching and testing for vulnerabilities using frontier AI models again, because we know the attackers are, is absolutely necessary. Again, something that we help our customers with. And then when we're talking about if an identity is compromised, that attacker can often fly under the radar of traditional security tools because as I was talking about in the opening, they're often acting like a trusted user, and the speed is so fast that even an AI driven security operations team isn't going to find an anomaly in that actor's movements until the very end of the attack chain. And so that is why the AI-driven security teams, the virtual patching, understanding vulnerabilities have to go hand in hand with identity security.

14:14 Because if you can truly secure all the identities, human, machine and AI, you are going to stop the attacker before they even need to be detected and responded to. So that's where identity security comes in.

Jon Krohn: 14:27 All right. Yeah, let's dig more into that all important identity security. You've argued that the line between privileged users and ordinary users completely disappears when we're in this now machine speed, agentic AI environment where being able to enumerate all of the security vulnerabilities costs nothing and takes very little time. Why has that line between privileged and ordinary users disappeared?



Amy Blackshaw: 14:56

Yeah, it's a really important point because in the security industry for a long time we've sorted into tiers this idea that privileged users are one thing, most likely an IT admin, and they're ordinary users. That used to work in a different world. When we were on-premise, we didn't have people accessing cloud and we didn't have developers working at the speed that they work at today, and we absolutely didn't have this AI foundation of threats changing the landscape. And so we have moved away from this idea that privilege only exists for IT admins. And the interesting thing is that in the old days, IT admins had to use tools like privileged access management that they would sort of bear the friction of using. And so we're talking about 20, 15 years ago, IT admins were put in this bubble and said, if we protect them, we're protected.

15:57

What's happened now is that every user, whether you're a developer, a data scientist, whether you are working in cloud infrastructure or if you're just a business user who has access to customer data or financial data or code base, you slip in and out of privilege every day. And so that is this idea that you need to put the right level of privilege control, commiserate with the risk, and commiserate with the level of friction that the user will bear. So this has completely changed the way that the privileged access management market and technology has shifted, and we also need to apply them to human and machine. And what's been great about this journey from CyberArk to Idera is that we founded privileged access management. We've reinvented ourselves. We now, of course, have traditional privileged access management controls, but are now having modern privileged access management controls like zero standing privilege and just-in-time access.

Jon Krohn: 16:58

Yes. And we will get to zero-standing privileges and just-in-time access shortly in this episode. Before we get there, I want to know about something called the trust window and why we need to focus on limiting the trust



window instead of just focusing on keeping attackers out of our systems entirely.

Amy Blackshaw: 17:16

Yeah, it is this idea that if every user is privileged or slips in and out of privilege during their day, the way that you secure that identity and that privilege has to be around how long the privilege and the entitlement lasts and how far it reaches. So that's the trust window. How long does that privilege access need to be granted for and how quickly can we take it away? So think about every piece of access as a key, and every key has two properties, how long it works and how many doors it opens. So a key that works forever and opens everything is your maximum exposure. It was that key to the safe on the door that we were just talking about in the hugging face incident. A key that works for 10 minutes and opens one door has a very small exposure rate. So most organizations are carrying thousands of forever keys that have long been forgotten and attackers now with AI can find them.

18:25

And that's why we focus on shortening that time window than just simply keeping attackers out. So we know attackers are going to get in, but if we can change that trust window to be the smallest amount as possible, even if the attacker gets that key, it's going to stop them in their track because it's going to be useless to them because it's out of the range of usefulness.

Jon Krohn: 18:47

I gotcha. And it sounds like this is going to be important for the zero standing privileges and the just-in-time access, which we are going to still get to. But before we get there, I want to talk about agentic identities because I think that's another piece of the puzzle to understanding zero-standing privileges and just-in-time access in the current agentic AI environment. So this is a data science and AI engineering podcast. Most of our listeners build AI agents. From a cybersecurity perspective, how is an agentic identity different from a human user or from a traditional service account?



Amy Blackshaw: 19:26

Yeah, I love this topic. And it's because agentic identities are so fascinating. And like we talked about in the beginning, there's so much promise and there's so much opportunity that we all see with creating agents. And from an identity perspective, they're very unique. So if you think about it, a human identity is a person and we move at human speed and we have human judgment and we have human accountability and we will do things to better our goal, but we're limited sort of in our humanness. Now on the flip side of that, a service account is the opposite. It's static, it's set up once. It's usually set up with too much privilege, and it is left so that it could run for years and sometimes has a credential that has never been rotated. So very, very different, static versus ever moving and ever changing. Now, an agentic identity is sort of a strangely unsettling combination of the two from an identity security perspective, because it has an autonomy of a person.

20:38

It can make decisions, it can chain tools together, it can sometimes act on someone's behalf. It will elevate its own privileges to get to its goal. It will go around security controls if it can like a human, because none of us like to be stopped by security, but it's provisioned like a machine and it's actually code. And so there are the chances that it can be set up once and sort of act like a service account. And there can be thousands of them, thousands of them we know created overnight. So it carries human reach with machine sprawl, and sometimes it doesn't have governance. So everyone needs to think about reframing the way that they think about identity security when it comes to this idea of agentic identities.

Jon Krohn: 21:27

That is crystal clear. Let's talk about what a rogue agent, like you were just describing, looks like in practice. Maybe you can give us an example of an agent being subverted or drifting from its mission or leaking secrets through flawed reasoning.



Amy Blackshaw: 21:43

Yeah. So I'll talk about it in three ways. So first, as you're saying, it can be subverted. It can be fed a document or instruction that is poisoned or malicious, and because the agent is trusted, its access quietly can become the attacker's access. And there have already been real cases in the news this year where an agent's own token was lifted and inherited everything it could touch. So that's number one. Number two, it can drift. It can pursue its own goal in a way no one intended, quite like the hugging face intrusion, reaching into systems that were never in its scope simply because the path was open and nothing stopped it and it got it to its goal. Or it can leak through its own reasoning, do exactly what it was asked, and then in the process, move a secret somewhere it shouldn't because it has broad standing access and no reason to think twice.

22:42

So the main thread that runs across all of these is really the hopeful part. In every single one of those stories, the damage is bounded entirely by what the agent was allowed to do and for how long. So back to our analogy of a key, if the key only is alive, able to access something for 10 minutes and only one door to open, the same way an agent, if it has the right guardrails and only has access and privileges to where and when they need it, we're going to limit that ability to go rogue.

Jon Krohn: 23:15

It sounds like you might have actually already answered this next question, but you might have a little bit more to say about it. For people like my core audience, AI engineers, data scientists, what can they be doing differently when they grant an agent credentials or tool access? It sounds like obviously having a time limit, a trust window on that credential is key. Is there anything else?

Amy Blackshaw: 23:36

Yeah, you're a fast study. I love that. Yeah, it's exactly that. You don't give standing long lived credentials, but access for the length of the task that expires on it on its



own. So for the rest of the time, there's nothing for it to steal. You've got it. It's about the time window. It's about limiting privilege and setting just in time, zero standing privilege controls on agents as you would a human, hopefully.

Jon Krohn: 24:01

Yeah. So it sounds like we might have actually now kind of gotten an intuition around what just in time access is and what zero standing privileges are. It seems to me, I can try to explain it. You can tell me what I get wrong or what needs to be developed on further, but it seems clear to me that this idea of zero standing privileges means you don't grant a standing privileges ever. You never have a key that works indefinitely. And then the just in time access part means that you provide the access. The key only works just in time or just at the moment that it's needed. And then as soon as it is no longer needed, that access is taken away. Is that right?

Amy Blackshaw: 24:43

You got it 100%. I like to stay with zero standing privileges. There's zero privileges, zero permissions, zero entitlement that is left standing for an attacker to come and take advantage of. Because if there is a credential, a username and password, a token or a secret that is decoupled from the entitlement, then they can get the credential all day long. They can fish, they can come in and laterally move and take over a credential. But if there is zero entitlement, zero privilege tied to that credential at that time, it's a dead end for them. We've cut off their access. So you got it 100% right. And I think when you think about this in practice, so what does it mean for a user who has zero standing privileges? Is there going to be friction for a user because do they have to constantly ask for access?

25:42

And the trick is to grant the access and revoke the access automatically and driven by policy and risk. So if there's an engineer that needs access to AWS infrastructure and there's context and there is policy, when that engineer



needs that access, it can be automatically granted and when that session is closed, it disappears. So you're doing sort of both. You're enabling the user to be able to move as fast as they want, and then you're stopping the attacker because you are stripping that standing privilege that for far too long organizations have been using.

Jon Krohn: 26:23 Brilliant. Thank you for more clear explanations, Amy. All of this concept around zero standing privileges, just in time access, it seems to me like that could add friction. It would be convenient for me as a developer, as a data scientist, as anyone creating a system to just have a key that unlocks all the doors and works forever. That's super convenient. So everything that you're describing, just in time access, zero standing privileges, that must add friction, right? How do we minimize that friction?

Amy Blackshaw: 26:53 Yeah. And again, it goes back to how it is deployed. So if you have a system in which somebody has to constantly come and say, "Can I have the key?" When you have to use the restroom at a restaurant or a Starbucks and you have to go get the key, it's annoying. The idea is to be able to open that door or request that access just based on who you are and what you're trying to do, which is set by policy. So there are certain things that certain identities are going to do that are legitimate time and time again that can be set to be automatically granted and automatically revoked. So that way you aren't introducing that friction because I always say the best cybersecurity can be the worst cybersecurity because if nobody uses it, it doesn't matter how secure it is. So it has to be granted by policy and by risk level, and it has to be done in a way that is natively in line with the user's tools and how they would approach their day-to-day so you don't have that friction.

Jon Krohn: 27:52 Makes perfect sense. Looking ahead a little bit, LLMs, frontier models are only getting more and more capable. So you have said previously that defenders have three to



five months from them hearing this episode before autonomous exploitation becomes standard. So if my listener, if their organization starts a 90-day plan tomorrow to get ahead of this autonomous exploitation wave, what should they do first in that plan?

Amy Blackshaw: 28:23

Yeah. I think the countdown, what we know is that attackers are getting faster and that the window to exposure is just becoming smaller and smaller and that it becomes a harder problem to solve for defenders who have to get it right every single time, where the attacker just has to be right once. And so whether it's three months or five days, organizations need to get started today. And if they are going to get started today or tomorrow, the first move that I would like organizations to see is to take standing privilege off the board. And it is a journey that many organizations are on and have been on, but if you can take that standing privilege away from where it matters most, that elevate that just in time so an admin isn't lying around waiting to be found, that would be the one thing because you can move to zero standing privilege for all of your users, but first start with your admins.

29:32

If your admins are using standing static secrets or access, get them off. Get off of standing hard-coded credentials in code for machine identities, get off them, use ephemeral secrets so that way we can stop the attacker. So machines, agents, humans, starting with your admins, get off standing access, move to ZSP, and that would be a ginormous first step for organizations to take.

Jon Krohn: 30:01

Fantastic, Amy. We're going to have to watch closely as this autonomous exploitation becomes more and more standard in the coming months. And yeah, we'll have to have you on again to explain the latest exploits that our LLMs are getting up to out there in the world and how we can be combating them with solutions like Palo Alto Networks Idera. Before I let you go, Amy, I ask all of my



guests for a book recommendation. Do you have anything for us?

Amy Blackshaw: 30:30

Oh, yes. I love to read. I always have two books going, one fiction and one nonfiction. So I'll give you a good one in each category. So nonfiction, I always love to recommend *The Happiness Advantage*. In the world of cybersecurity and spending a lot of time thinking about attackers, it's nice to think about how happiness can be advantaged in the business world. And it is something that I subscribe to 100%. And it's not about being Pollyanna at all. It's about being able to identify opportunities to leverage happiness as an advantage. So check that one out. And then for fiction, I will say *Theo of Golden*. I know that it's one of the highest recommended on good reads. I have given this book to so many friends and family. It is just such a good read, *Theo of Golden*.

Jon Krohn: 31:27

Theo of Golden?

Amy Blackshaw: 31:28

Yeah, you got it.

Jon Krohn: 31:29

Wow. Okay, cool. I've been looking for a good fiction recommendation recently.

Amy Blackshaw: 31:34

Oh, read it and let me know. I'm telling you. It's a book you'll throw across the room out of sadness and frustration, and then you'll go run and pick it up to finish it. It's great.

Jon Krohn: 31:45

Okay. All right. Looking forward to that experience. Amy, final thing before you go, how can listeners follow you or Palo Alto Networks after this episode?

Amy Blackshaw: 31:54

Best place is probably LinkedIn. Check out the Palo Alto Networks handle. There's also an idea by Palo Alto Networks handle and myself. And yeah, you'd be able to find different blogs and pieces and content that we have out around spaces. I think I have a piece recently written called *You Don't Have Three Years*, which is on the topic



we talked about a lot. That's a good place to start is LinkedIn.

- Jon Krohn: 32:21 Love it, Amy. Thank you so much for taking the time out of your valuable day. All of your time must be so in demand, but you gave us such a brilliant episode. Our audience will appreciate it for sure. Thanks and hope to have you on again sometime soon.
- Amy Blackshaw: 32:35 Thanks so much, Jon. I've enjoyed it.
- Jon Krohn: 32:37 What an exceptional guest Amy Blackshaw was. Wow. I love that episode. In it, she covered how nearly every breach comes down to an attacker getting hold of an identity they shouldn't have, why Frontier AI hasn't invented new attacks, but has changed the economics since the expensive expert thinking behind chaining vulnerabilities together is now instant, tireless, and cheap. And she talked about how the first move for any organization is taking standing privilege off the board, starting with admins and with hard coded credentials in machine code. I hope you enjoyed the conversation. To be sure not to miss any of our exciting upcoming episodes, subscribe to this podcast. But most importantly, I hope you'll just keep on listening. Until next time, keep on rocking it out there and I'm looking forward to enjoying another round of this SuperData Science podcast with you very soon.