CTL-CISO-Phil-Stafford

Mon, Jun 22, 2026 5:53AM 13:12

SUMMARY KEYWORDS

AI governance, legal accountability, fractional identity, confused deputy, MCP servers, security lifecycle, supply chain attacks, AI S-Bomb, due diligence, agent permissions, sub agents, cloud apps, security theater, AI security, best practices.

SPEAKERS

Phil Stafford, Jo Peterson

J **Jo Peterson** 00:07

Hey everyone, thank you so much for joining this episode of Clear Tech Loop. I'm Joe Peterson, I'm the CIO of Clarify 360 and the chief analyst at Clear Tech Research. And today I've got mr. Phil Stafford here with me. Hi, Phil.

P **Phil Stafford** 00:22

Hi, how are you doing?

J **Jo Peterson** 00:23

Good, thank you for joining today.

P **Phil Stafford** 00:27

Glad to be here

J Jo Peterson 00:28

Yes, nice to have you here. So, Phil is an AI security architect and researcher. He advises founders and companies on AI security and cyber security foundations, including best practices and security customization schemes. Additionally, he shepherds SMBs and startups through AI awareness and implementation to self-sufficiency. And who doesn't want to be self-sufficient? Everybody does. So, as normal, we have three questions. And let me get going with the first one, Phil. How do we operationalize AI governance? That's the first part, and then who is legally accountable when an AI agent makes an unauthorized decision.

P Phil Stafford 01:14

So, the first part about governance, the first step is measurement. You need to know what your agents are doing. A lot of organizations will just start writing policies for best case scenarios, or what they think it should be going on. That's not really governance, that's just liability theater. So, first you have to make sure you're measuring what agents are actually doing in your environment. As far as accountability, the problem is that legal system really hasn't given us answers yet. We're still working all of that out, but in the meantime, what organizations can do is make sure that they can prove that they exercise due diligence before they've deployed anything. Then you have evidence and you can show that you actually did something.

J Jo Peterson 01:58

I love that, and I love the term legal theater. I think you should coin that. That's great, you know. It's, I mean, it's such a dicey environment, and I would hate for the poor system to get blamed for an AI agent that made a bad decision, right? Because whose fault should it be? Not, you know, it's.. it seems that we're. is that is where the arrow goes these days.

P Phil Stafford 02:23

Yeah, I think that it's really about whose authority the agent was was deployed under, not necessarily pushing the person who pushes the button, but the person who signed off on it.

J Jo Peterson 02:34

Right, right. And you're right, the legal system is just sorting some of this out. It seems that legal is always a bit behind tech,

P Phil Stafford 02:42

always

J Jo Peterson 02:43

just the way it works, right? I just, I like this. I just heard MCP referred to as the confused deputy. How do we prevent agents from executing actions that the user should not be allowed to perform?

P Phil Stafford 03:02

So the confused deputy issue arises when you have an agent that's deployed, and it, it, the system sees your agent as you, it inherits all of your permissions, so in the course of doing its task, it also doesn't know what it shouldn't do, it just knows what it can do, and it has your full permission set, so it does anything it can do to get to its goal, which may not necessarily be the right thing. It's like if I gave an intern my badge to get into the office, but my badge also lets them into the server room.

J Jo Peterson 03:38

Yeah,

P Phil Stafford 03:39

that's a problem. So part of the issue is really just scoping down permissions for agents, maybe with fractional identities, so that the system then sees Phil's agent, not Phil.

J Jo Peterson 03:54

Yeah, so take, get, let's stand there a second, because I'm not sure that everybody understands the concept of a fractional identity. What is that? A

P Phil Stafford 04:05

fractional identity, so normal normal identity is then the system recognizes you as you with all of your permissions and authorizations. A fractional one ties that to you, but is more of your delegate, so an agent is still you as far as the system is concerned about upward stream, like accountability, things like that, but it can only do a limited set of your permissions, so it's a subset of your permissions.

J Jo Peterson 04:35

Okay, so, so then let me tease that out a minute, because I think that's fascinating with this fractional identity, is it then a sub identity of mine? How are we baselining and accounting for that identity?

P Phil Stafford 04:52

You've got it right. It really is kind of a sub identity we're using, you know, in this case, you. The accountable person, and you're the one with permissions, but it only has a small amount of your permission, so the system now constrains that agent even tighter, so that now it, what it can do, is limited to what it should

J Jo Peterson 05:16

do. Okay, so two other questions come to me from that. The first is the idea of duration of permissions. Does this subset of my identity have my sort of like I could log into something 24/7 How about this subset? How do we rationalize for the amount of time that that subset of my permissions is allowed to access things,

P Phil Stafford 05:45

it's the same kind of thing. That's what fractional identity helps with, is that it now has limited time. You can have limited permissions, you can have limited time, any sort of behavior that you would normally do, it now is constrained. So maybe you're working from nine to five, but this agent's only going to run at 11, only going to run right in the morning, that's it. If it runs at any other time, the system flags it, even if you are supposed to be there, right? Agent wasn't supposed to be on, and so it notices it's we've got the tooling that we're working on. There's a lot of tooling in this space right now, yeah, but everything's new. That's the thing about this field. Every day, there's something else.

J Jo Peterson 06:25

Yeah, literally

P Phil Stafford 06:26

every day.

J Jo Peterson 06:26

Yes, it, yeah, it feels like, oh, right. And then let me stand here a second, because so this subset of my identity, this partial agent, or this, this partial identity of mine, can it spend spin up an ephemeral agent on its own?

P Phil Stafford 06:48

That depends on if you give it the permissions to do that, you know. There's tooling that allows you to do that. Sub agents, Claude has that right now, where it also spin off sub agents. If you give it permission, absolutely, it could kind of

J Jo Peterson 07:03

why I asked the question. Yeah,

P Phil Stafford 07:05

yeah, and then you have the same, the same issue, where the sub agent, if you don't have any more controls, would inherit all the permissions of your agent, so maybe you want to give it a smaller fraction, that's why fractional is the word to work on.

J Jo Peterson 07:21

Yeah, that's slice

P Phil Stafford 07:23

that pie even smaller. Let your agent go put a bunch of sub-agents out in the world, but with even more limited permissions, so that they're only doing the thing they need to be doing, and that's all because they'll try, they'll try really hard to do whatever it is you set them out to do, and they just don't know what they shouldn't do,

J Jo Peterson 07:44

what they shouldn't do, right? Okay, another thing that drives me a little bit crazy, I know we need it, I know it's a necessary evil, is MCP servers, and it drives me a little crazy because it feels wild west, just to me, I don't know, so let me stand still there. Any, any personal grudges against MCP servers that you'd like to share? Do you love them? Do you hate them? Are they, are they, are they met, like oatmeal to you? Like, why?

P Phil Stafford 08:17

Well, I don't. I wouldn't say that I have a personal grudge against them, but I do, in the way that we deploy them, and this is this is this security, the security lifecycle in a nutshell. For the last 50 years, we make a thing that works, and then we try to secure

J Jo Peterson 08:34

it. Yes,

P Phil Stafford 08:35

which is, of course, it's how you would do it, because why wouldn't you? But sometimes we're a little too reckless. One of the things that MCP was sold to us as, and I think it's a really apt parallel, is that it's like the USB for AI.

J Jo Peterson 08:52

Oh, that's good,

P Phil Stafford 08:54

which is great. I think it's great. That's exactly what it does. It allows you to talk to other tools universally, but it's also the USB for AI. You would not pick up a USB stick in your parking lot and put it into your enterprise environment. That's what people are doing right now. That's what's happening. So it's really, it's a matter of knowing what it is that you're using, knowing which servers are good, which servers are bad? Having an approved list in your enterprise environment, and then not allowing other ones to run

J Jo Peterson 09:27

well. Let's talk about the not allowing other ones to run thing, because that seems like a great idea, but what's happening is everybody's got an MCP server, and so now the security team is tasked with authenticating and trying to figure out security for third party MCP servers. How do we do that?

P Phil Stafford 09:53

Well, it's very similar to when we all had cloud apps, we all had that. Shadow, it's going on, where you'd have users just installing whatever they could to get their job done, very similar to agents, they're going to do whatever they can to get their job done, not necessarily what they should do the right way, right. Okay, so if we're scoping everything down to an allow list of MCP servers, and you have detection tools in your environment to know what else is being run. There's a call out to an MCP server in, you know, a country that you know you don't, you don't have anything in,

J Jo Peterson 10:31

right?

P Phil Stafford 10:31

You can flag that, and you can stop that. That's a simple, simple way of doing that.

J Jo Peterson 10:35

Yeah,

P Phil Stafford 10:35

but a lot of this is supply chain too, just, just like recently, you know, the supply chain attacks are all the rage now. We solve this for security or for software, you know. We did desk bombs, we did co-signing dependencies.

J Jo Peterson 10:51

Yeah,

P Phil Stafford 10:51

we just do the same thing for MCP servers. Know what your MCP server is doing, what it's calling, what it needs, and then where in that chain might be weakest,

J Jo Peterson 11:03

so I mean, I think you're a fortune teller, because I think we're going to start seeing S bombish type software for MCP servers. I think that's what we're going to start seeing come out, because how could we not? Because it's

P Phil Stafford 11:17

absolutely

J Jo Peterson 11:18

right, it's too hard to figure out otherwise. I mean, you know, it's kind of like in my head going, well, how do we blacklist

P Phil Stafford 11:28

it? Right, it's, it's, it, we have the controls already.

J Jo Peterson 11:32

Yeah,

P Phil Stafford 11:32

it's not magic, it just means thinking about it the same way we always have and reapplying what's different, so it really is more like a software supply chain than anything else. I'm working on the work stream for the Coalition for Secure AI's S Bom, actually AI S Bomb,

J Jo Peterson 11:54

that's so cool. Yeah,

P Phil Stafford 11:56

I know, and that's the kind of things that we work on, is really like building things out the exact same way using the exact same formats. What does this MCP server need? What does it do? What do the dependencies need, and what do they do? Because that's that's where you, that's where we're getting hit. People are sneaking in something up their supply chain.

J Jo Peterson 12:18

Yeah, and now, and now on top of it we find out our AI doesn't like us through Mult Book, they don't like us, even

P Phil Stafford 12:30

I don't know about that, I think I think Malt Book was AI, AI knows how to pretend to be Reddit.

J Jo Peterson 12:38

Oh, nice. Okay, that's cute. All right. Well, you know what? We had another guest, Gerald, recommend you, and he said you were smart and fun, and he was right. So, I'm so glad that you took some time out of your day to come visit, and y'all, thank you for taking time to visit with us as well. Feel nice chatting with you.

P Phil Stafford 13:00

Thank you so much. Great to be here.