FEDERAL TECH PODCAST

Ep. 142 Applying AI to Threat Intelligence

00:03
Hey, this is John Gilroy here. Out of all the interviews I've done over the years, I never thought I'd be talking about the Tower of Babel hit the music Mannie. Welcome to the federal tech podcast, a podcast that connects you to federal Technology Leaders. My name is John Gilroy, and I will be your moderator. No, this isn't a Bible study, we're going to talk about a company called Babel Street. And we're gonna talk about how it helps you make your systems faster and more secure in all different topics. And our guest today is John Weaver. He's the Chief Strategy Officer of Babel street. So John, unpack this Babel street. What's going on here? What do you guys do?

00:44
John, thanks for having me on. That's a great intro. And I appreciate it a babel is our core is think about the idea that it's not a not a Bible study. But it does come back to why you have explanations and why there's so many languages around the world, the challenge you have and think about the global economy and global information space. If you think about just being able to understand somebody who's typing something in English and wants to search in Spanish, or search in Chinese or Arabic, or Tagalog, and find out information that's relevant to them for making their decisions process, that's what we really help them do. So how do you take lots of information? Even if it's an English, imagine a chemical compound or a bill of materials for software labels? How do we transfer all that into something that is understandable to the to the user or the system? How do you turn that into something that's actionable? And that you can you can go further with that understanding from Babel. And so that's why we borrowed the name from Babel Street. At one point, it was also the embassy row in Cairo. And so as the world's leaders and what's companies countries met, on that particular road, and how do you decipher it and work things out together? Well,

01:47
I don't want to make sense. I mean, a malicious actor in let's say, the global south is not going to be communicating to his colleagues in English, they don't speak English. And so it would make sense to have this kind of sensitivity. So I imagine the origins your company may have been with the intelligence community, is that correct? A

02:02
lot of our original work was within the the intelligence community or the national security community within counterterrorism mission. So how do you understand if someone is giving instructions on on how to construct a improvised explosive device and they use the term que no, three, but it's written in Arabic. And that happens to be happening. Nitrate was a precursor. And so as somebody who grew up in North Carolina, I speak English and I sometimes speak lewd other languages, but I don't know what that looks like if it's written in another language all the time. And so can I use that as a native English speaker, find the information that's relevant to me to be able to identify actors and actions are taking place?

02:38
Well, you came to the right place, my daughter's speak some Herrick, and Carol Wanda, and her husband speaks Arabic. And I know a little bit about Portuguese, my wife's a Latin teacher. So I understand some of the challenges in language. It's not all one, two, you can't just go to Google and say it's raining outside and type and go word for word for word. So there's, there's some subtlety and nuance in this communication isn't there

03:00
100%. And where your daughter, I heard a couple times that your daughter has done some work in Africa. And if you think about the sheer number of languages President present within a single country, they're being able to take the context of what the conversations are taking place. And being able to take that context to get the correct translation. Google and Bing have great tools to translate content at volume. But when you do that translation machine, even as advanced as they are the nuances often missed. And so the other thing you don't want to do as a public sector organization, for example, is translate the internet, that is cost prohibitive. It's time prohibitive. And so being able to find the things that you need within the language within the context that to solve the challenge that the public sector organization or private sector organizations looking at, and bring that back, that's really where we come into play. I

03:46
love that phrase. What are you going to do today? John, I'm going to translate Google. Oh, good. And what are you to buy for lunch? And so I think this is a question. There's so much information everywhere. And so So John, how can your company help my listeners, you know, unlock the insights that matter of us? Absolutely.

04:03
We look at it from a number of different language areas, from how we solve problems. So we really help public sector and private sector organizations understand the risks that they're faced, it might be risks affiliated with individuals organizations, can I do bank? Can I bank with that individual? Is this person on a sanction, list? Should they be granted access into the United States on the on a watch list? Can I do business with them? Are they on? Do their beneficial owners have some challenges that might make it untrustworthy or violation of sanctions? do I how do I actually just help organizations get registered within a business database or something like healthcare.gov? So how do you do those things in a way that is efficient understand about the risk and decisions that have to be made there? We also we also help organizations with with threat intelligence and so what are the actions that are going to take place that would have adverse impacts either on my people I organizations or my my processes, on How do we do that in a timely fashion? If you think about global sporting events, whether it's a World Cup or the Olympics or the NCAA Tournaments, you know, how do we help them? The folks that are managing those events must stay abreast of what's happening so that they act appropriately. And then finally, if you're dealing with people, how do I ensure that ordinary people in organizations? How do I ensure that information that is proprietary to organizations retained inside it isn't linked inadvertently or and or intentionally? How do I make sure that insider risk is appropriately managed, you know, that's there's a risk side of that and a threat side of that. Sometimes it's just inadvertent, I casually put information out online, that maybe discloses information that shouldn't be there, or I am intentional activities that will say I'm trying to harm or otherwise cause cousin incidents with that organization. So we help organizations put that together, we help

them meet where it is. And then we bring that in, so they can take those actions on it, whether it's on a data side, a name matching side, or even an access to information side and put that within their workflows. Don

05:57

Weaver, we're gonna give you an English question, because that's your proficient language. And this is a simple question. I deal with a lot of satellite in space, folks. And when they talk about situational awareness, they're talking about a satellite coming at them and not being able to avoid it. And this is this increasingly a problem, it's gonna be 30,000 satellites up there in the next few years. However, when I was at your website, I noticed that you use situational awareness maybe in a different way. So what does that mean to you and your organization?

06:23

Absolutely. And I'll take it back even to the satellite industry. So if you think about a company producing satellites, for example, that company needs to understand what are what are impacts to their supply chain, the components they use to put that together? What are the threats that might come from other nations or other companies where they're trying to exploit their key intellectual property? Where are they trying to exploit information vulnerabilities within the satellite industry, as well as satellite is no longer something that government just puts together, and it's all done in one place? It's hundreds of 1000s of components long. And so can you identify vulnerabilities that adversaries may see it may seek to, to exploit and even getting something as simple as how do you deconflict the number of satellites that are that are in space. So giving an understanding about these are incidents that are being reported about them. These are things around launch facilities, these might be protests that take place, or frankly, as I said, there might be adversaries that are there, or just here is information that's publicly available about that entire spectrum from design to deployment. And so we try to get that holistic set of information to companies that are trying to are and organizations that are trying to make decisions about what to do next, do I increase my security posture? Can I source my components from a different place that's more reliable. All of those are information pieces that we really try to bring together. And and the last point I'd even make is even understanding the vernacular about the satellite industry. What are the acronyms that come into play? What are the slang that come into play? And if I get content from my fliers, it's coming from all over the world, and I need to manage that as a knowledge set? How do I appropriately do that? That's where Babel street really comes into play and creating that different view of situational awareness that is dependent upon that organization and where their risks and perceptions are.

08:06

Risks and perceptions. That's the phrase of pace. I was at your website. Got a couple articles there. I think we have a whole article there on risk conference camp

08:14

to you. Absolutely. Thanks, John. Roos confidence gap is something that we've been working with clients and partners for for a little while on, you talk a lot on your show about how it's not hard to see how data is really overwhelming decision makers and policymakers. There's no shortage of it, you can get from social channels, or news channels on internal and external resources, there's more data that can be consumed by a single analyst. And that's even humanly possible no matter how large the budgets are. And what that happens is

when you're when you have that much data, and you don't have more people necessarily, and even you don't have necessarily mature AI ml processes to help you analyze it, you create a gap and your confidence to be able to meet the risk your nation fate organization faces with the resources that you have at hand. Every company and every organization has to do the best they can. Sometimes that's a manual process, what manual process means is you will not do that consistently all the time, you will not necessarily get to the 17 page of a search results to find out the answer, you certainly won't be able to do that in another language with proficiency and confidence. And so what we try to do is help close that gap. And the ability to tame the hear of huge amounts of information that's out there. And then do so in a way that is as consumable and usable by organizations trying to manage that gap. That risk could be overcrowding of the space, the satellite network, it could be insider threat, it could be external threat. It could be supply chain, risks, regulation, all those things are things they can be provided across the horizontal that we can help them solve.

09:45

And I think the article I'm referring to is called what is the risk conference gap? And what I will do in the show notes for this episode, I'll put a link there so the listeners can just hit hit mute on their phone and go and read it's interesting articles got some time Next about zettabytes and everything else. So it's kind of a good one. Um, in researching this interview, I discovered that you acquired a company called vertical knowledge. And just trying to figure out how that fits in your product offering and and how's that have any benefit to my federal listeners vertical knowledge?

10:17

Absolutely. Thank you for the question. So we merged with vertical knowledge back in January of this year, they have been a partner for a little while before that. One of the things that vertical knowledge, one of the many things that vertical knowledge has done that attracted us and felt like it was really complemented for our for our customers, was their ability to go out and find information. And bring that information back in a really structured way, that's a lot easier, that's that, that can be complimentary for their solutions. And so a good example of what what they do, for example, excuse me, a good example for what they do is they can go out and find pricing information. So if you're a commercial company, you're trying to find prices and understand what the competitive market looks like. And you want to invest further in a particular market, they were able to go out and find that content, regardless of where it is. And also from the point of view that it takes place. And so if I searched, we're just talking before the show, I just got back from vacation, I was in southern Utah, I'm searching for something that's actually relevant here. In the DC area, my results were from Southern Utah, not from the DC area. And so being able to understand that local Point of Presence and how search brings back information that is relevant there. It has great uses for helping Oregon public sector organizations understand and manage risk, because you get that local perspective on information. But it also helps make great decisions, a lot of investment organizations will use those and say, Hey, am I investing in the right company within the satellite industry? Am I making the right choices from a from a supply perspective, or from a market perspective that would go in there. And so they really bring a holistic math mechanism to help bring that content together. When we marry that rich content with rich analytics and rich AI and that language understanding of language, you're able to, we're able to serve more clients across a broader set of use cases that encompass both the public sector and the private sector.

12:07

Okay, so John Weaver is in Utah, I'm going to Utah in August, what a beautiful place to be. And he goes to his company site, he's sites to download the white paper, or some document now. Oh, are you buddy? Are you John Gilroy, John Weaver, the man on the moon, who's this guy from Utah, no one's ever logged in. I mean, the whole idea of identity management fits into this concept of, you know, who are you and multilanguage systems? And, and and a lot. So you folks do Identity Management tour work in that general area, don't you? We

12:37

support a lot of clients and partners who do do tie identity management and leveraging in different ways. And I'll give a couple of examples. One, for example, was when the Affordable Care Act went into place, there was the requirements for the registration. So you needed to the government requirements included, how do you match birth certificates and death certificates. So you get the right person, you think about the multiple government databases that were that were in place, they needed to get all that together. So you get a single essentially login for for healthcare.gov. The client we were working with was doing this in a manual fashion hired 1000s of workers to be able to do that. And then when they came to us and we worked with them on is how do you actually create an automated process that matches the right name to those write records. And so when you start from that, that more consulted, consistent single set of proof that it helps you do that easier. For example, I flew yesterday, my name was concatenated on my boarding ticket, if I have the airline flew into the hotel I normally stay in and my passport, they're all slightly various, they shouldn't be but how do you make those all together so that when the when we actually got to check in, it's actually me checking in on me having been screened, and that me getting the right account, I just want to know I'm in the right seat in the right hotel room. And so that's really what we can help organizations do. And in that case, we worked with a partner who built the Indian system and was supporting that requirement, we help them do that in a way that save them in a single quarter, 70 days of time. And so that's a huge cost savings, if you can just reduce by percentage point, sometimes the amount of manual interventions are required and let humans do what humans do really well, which is kind of recognize these discrete patterns. So that's an example of what one area where we get into identity management. Another area might be, we talked a little bit about elastic and observability. And how they think about the cyber threat. If you take the kind of ones and zeros of observability person logging in a certain time. They have this particular pattern and behavior. You know, you kind of get indicators within insider risk or insider threat that might be there. What if you took that one step farther and further and said, What publicly available information about that user might say, when you hit a threshold, maybe I need to go out and look and see if there's any concern out there? Is there an inadvertent leak that might be taking place because of behavior? Or is there an intentional threat that we might have to be realistic? And so it becomes another facet of identity management? How do you tie the digital profile of an organization or an individual back to what you might have? Basically

14:59

that's interesting. To when, when he talked about humans, I flipped it and talked a lot about nonhumans and robotic process automation and identities and non human identities and, and it's one thing okay. John Weaver, give me your ID Okay, great. What about a process automation, which with the speed that systems or flood

was today? And so one of your blog posts, maybe this is from vertical, they talked about entity resolution sounds like a sci fi issue a chapter six entity resolution. See folks do that too, huh? Absolutely.

15:32

We play it play a role in putting that together, you know, the idea of a digital persona of a person or organization? How many different ways as a name get represented? How many different ways is the address get represented? What are the other elements that time that from a digital perspective that can tie back into that identity? Say this is actually the John Weaver? Who was flying? Who did these things are this is the Babel Street who sold this product to this organization? And so how do you take started the name use information that you find in other sources to make sure it's the name is the right person, it's not John Weaver that's in California and stood up and said, it's John Weaver that's in Washington, DC. And so we play a role in that disambiguation that comes to around organizations and people and concepts. And so using that being able to link those together, makes it so that you can automate more of these processes. There's no no system is perfect. But if we can reduce the number of times that a human has to intervene, and that that means that you can, you can have more applications approved by banks in a faster way, you can reduce the backlog and the time it takes to secure a visa or to register trans international transactions. All those things are often about how do we just improve the user and customer experience that either the public sector organization has or private sector has reduced the number of people that actually have to have a human review. And then when you have areas that might be yellow, or concerning use humans for those where it's more appropriate. But all in all, reduce that time it takes to get an action done, and reduce the time it takes to a customer get access to the service, or the the investigator to reach the end result from a fraud investigation or other sort of criminal investigation?

17:19

What's interesting about the blog post on your site, I've gone to YouTube, and then we'll research on you is that you just don't you know, just look inside the continental United States. All the answers are there. Well, just look, we talked to this go to Google Google has has all the answers really? Oh, yeah, really, you know, you guys go into the dark web to go, Hey, what's going on here? Maybe there's a guy impersonating John Weaver over here. And let's say a different country, different language, even you know, I mean, you so you're drawing in all kinds of sources, from the open sources and getting this information saying, yeah, that Weaver guy in Utah? Yeah, he's the same one. But the dark woods Pirates of your offering, isn't it?

17:57

100%? In order to really answer the question about does a digital profile line up with a physical profile, that intersection of reality and digital space, you've got to be able to pull in a lot of different types of sources, the dark web is, is a place definitely where people think they can hide, and they don't really do as good of a job at that. It's amazing how often someone will use a user name on a public site that you might find on social media. And that same name shows up on some variation to that name shows up on the dark web, people that are on some public social media sites that may be involved in fitness. Trafficking, for example, will list their dark web credentials or their other credentials on those public sites. And then you have to go find them. And so being able to, to go out and allow organizations to make queries to go out to those sites. Without them there are people personally having to go out and do that provide security for the organization, but also completeness for

the investigation. It's not just about dark web, and deep web, and then kind of the rock broad range of sites. It's where information changes, the location, the conversations take place changes all the time. Ask a child who's in a person who's in college now how much they use Facebook versus Instagram. What do they use blog posts? Where do they do their reviews like that dynamic nature of both the places where the conversations and the information takes place? And the language that's used to describe it, not just English, or French, or German or Spanish, but what are the actual terms and how those change. So where we come to play and help our clients is as these locations change as the language changes or vernacular changes, and you need that fit in your workflow, we either provide a full end end solution, that's a SaaS application, you can log into and use it or as an API that plugs into their existing workflows. That's where we really tried to help them solve it from that perspective. Bring lots of data, lots of understanding single points of entry so they can better solve the problem as they understand it.

19:55

If you take a guy like Yeol, he talks about looking at the market writ large and He says, Well, you know, Facebook, maybe trending older. And so maybe the malicious actress or maybe under 35 may not have anything with Facebook, maybe on some social media platform, is it Instagram? Is that where it may be the correct answer today? And then John, Weaver six months from now, what's the different answer? And so, in, by the way, it's 24 hours a day, seven days a week, it's changing all the time. And, and, and Tay, I know people that really avoid email, and they may just use Slack and communicate with it. And so there's so many channels out there, it's hard to get a global view on it. We talked about satellites and looking down, well, that's changing one paths and other paths, and it's changing too fast. But isn't it I mean, so many sources. Now, that

20:43

is part of what causes the risk confidence gap is that just understanding of where the data exchange takes place, and that and the format of the data exchange, you know, where where we really tried to bring it in is that you have structured and unstructured datasets, you have publicly available content, that can be anything from court records, buying information, social content, so being able to pull that together and resolve it back to that right organization, or right person and right action. That's really where we, we help our clients the most, helping them understand the dynamic nature of that shift. If you think about somebody now who might have used Google, or we would go to Google and look automatically up to something, if you're, they're going someplace, they may 1, go to Instagram and search there, it's not something that would have occurred to me just based on you know, our age. And so being able to we work with our clients don't understand that, but they also inform us. And that's really, when I as a reformed product manager, I was one of the first things I did at Babel Street was our end product was, I love hearing where customers are starting to understand where the challenges are, hey, I'm seeing this evolution, how do you help us go get that and where we proactively work with that client to make sure that we're bringing in not only the content that we know is out there, and that we found, but listening to our clients and our partners, to say these are places to bring it and this is where it's evolving to. And so it becomes a really great relationship when we can learn from each other and improve the process with each other to make sure that when they have that challenge to find the find the bad person doing the bad thing, or make sure that the good folks can do what they need to do not trying to put it in the kind of polarizing view. But how do you get those right actions in the right people in the right place? We like to partner with them on I'm

22:20
glad you use word challenge because those of you bring that up, looking down the road at the challenge is just in all the things to handle language variations, and, and I think people think, you know, in fact, my kids and I were at Easter, were talking about how language changes in Noam Chomsky and how what a word may mean today may switch and change. And he talked about acronyms, and you know, colloquial phrases that change. And and it's one thing it's difficult to keep up with English. But what if it's in a non indo European language? If it's in Korean, and it's a an acronym, it's changing? It's this, it seems it's an impossible. So what challenges are ahead for Babel Street in the next five years in regards to this understanding linguists, that

22:58
is something that we embrace, and is really part of our ethos, here, the understanding of language and the evolution is really why we're here. You know, we we started with this, how do you recognize a name that might be matched inimical, or patronymic? will drive for the mothers or the fathers, right? But then you think about well, now I need to do that in a different script. And what does that what is that same kind of honorific if you're an engineer or a doctor? What was that come back to do? Where we focus the application of things like artificial intelligence and machine learning is it's not it's not every word that we use. But it's a big core part of how we discover it. We build models that help evolve over time, we can train them more quickly than a lot of organizations, can we make them contextually and linguistically relevant? We train it on data that is relevant to that topic. And then we can bring that back. And so as language evolves, that's something we actually built from our core bedrock of technology stack is, each of those things can be both explained. And why you get a response back, Why does John Weaver actually match? If John Weaver shows up in a Japanese registry? If I'm coming to visit Tokyo? Why does it match? And then how, as that thing changes over of that particular rule changes over time? How can I make sure that the organization continues to evolve to meet that devolving change, and so we work with them to train new models, and have new ways that explain that information as it comes back. And so we are not static, just like languages. And that's really what we build from. And so we see that as organizations continue to face the same dilution of information in the various language and representations that shows up that we give them the tools, whether it's the entire solution or parts of the solution that help them meet that challenge. And so we want to be there with him as partner for that.

24:39
may change change change. Yeah, I was when I first read Noam Chomsky, I thought about and so yes, it does. It's true. It does change all the time. It's not something that printed a book may not be able to handle anymore. Well, this has been a real fun conversation here. You have been listening to the federal tech podcast with John Gilroy and he website I mentioned during this interview, I'll put in the show notes page length thank my guest John Weaver Chief Strategy Officer at Babel street thanks John

25:04
thank you john