#200 - Copywriting AI (with Mark Rasch)

[00:00:00]

End Intro

Hello, and welcome to another episode of CISO Tradecraft, the podcast that provides you with the information, knowledge, and wisdom to be a more effective cybersecurity leader. I am here today with Mark Rash, with whom we had another episode, if you go way, way back to episode number 49. And if you're wondering where in the world is G Mark, where are we, Mark?

We are outside of Dublin, Ireland for the computer conference called COSAC. Exactly. So we're card carrying members of COSAC. I've been coming here over a decade. Mark has been coming here even longer than I have. It's a fantastic opportunity. And I think I stuck that up on the LinkedIn feed. So if you're interested about it.

for next year anyway, cause this year's already started. and if, speaking of LinkedIn, if you haven't followed us on LinkedIn, please do so we've got a lot more than just podcasts for you for CISO [00:01:00] Tradecraft. We give you a good steady stream of high information, high signal, low noise to help you with your career.

And if you're watching us on a podcast channel, hopefully you like what you got. Give us five stars so other people will help find us. But today we're going to talk about, Artificial intelligence. Okay. I've done AIs. We've done a whole bunch of AIs, but I haven't done an AI with an attorney yet. And Mark, if you remember, is an attorney and we have some common backgrounds together in terms of cities and things like that.

It'll probably work its way out in the conversation, but I'm surprised I got this many words in before he did. But, Mark. Sure. And thanks G Mark. We're going to have Mark and G Mark, here today. And so one of the things that we're going to talk about today is a couple of things about artificial intelligence and the law.

First of all, we're going to talk about some of the legal issues associated with what artificial intelligence is, how it works, what it does, some of the copyright issues associated with artificial intelligence. We're going to talk a little bit about

AI generated pornography and AI [00:02:00] generated obscenity and the legal issues associated with that.

We're going to talk a little bit about who owns things that are created by AI, and what, whether you can disseminate artificially intelligence created materials as your own. and then some of the privacy issues associated with AI, both in the terms of data collection, data analytics, and data output.

So AI, like any new technology, challenges what we thought we already knew. And that's a very good point because we've had some form of AI around for a long time, but it's a generative AI and the recent incarnation of the being able to go ahead and make it generally available to the masses. I think OpenAI gets a lot of credit for that by democratizing, if you will, the concept of an AI engine.

But it's interesting, I was talking to one, Company the other day and I was asking them about what if you go ahead and you incorporate some AI tool And you really want to make sure it works correctly and they said We'll do a code review I snickered a little bit and says you can't [00:03:00] do a code review on AI because the AI might be perfect however comma the code review it needs to be on the data that you train it on because the model is only going to give you what it is That you feed into it.

And as I said before it's not garbage in garbage out It's Garbage in, garbage stays, and then it gets pregnant and gives birth to triplets. So from a legal perspective, Mark, what is AI? Is there actually a threshold beyond which you can call something AI or is it just pure marketing? It is. Mostly marketing.

Now, AI has various characteristics and predecessors. So when we used to talk about machine learning, AI is machine learning. But it's also heuristics, it is a whole bunch of other things. It is basically teaching computers how to mimic human reasoning and how to mimic human communication. So we've been using forms of AI for decades, but what we're now doing is using AI and [00:04:00] the generative AI to be a substitute for things that we've been doing.

So within AI, there's two broad categories of AI. and these are directed AI and undirected AI. And I'll give you an example of a directed AI. Let's say we wanted to know, we want to work with a hospital or doctor to help them determine whether or not, something is cancerous. What a tumor is. And we can feed into a, a program, pictures of tumors, pictures of, things that are not tumors.

And we can have a doctor say, this was a tumor, this was a tumor, this was cancerous, this one wasn't, this was cancerous, this one wasn't. And what it will do is the AI program, as we have directed it, it will say, okay, I'm determining what the pattern is between what is cancer and what isn't.

You've determined what's cancer and what's not. I'm going to look for the pattern and now you can take an unknown photograph and the AI program will say, [00:05:00] based on what you told me before, this is or is not. The second thing you can do is to say the same thing with anything where you've told it what the parameters are and you've told it what the outputs are and it tells you whether or not it meets those criteria.

Now the advantage of directed AI is that it is a combination of machine learning and human learning. The humanist determines something and the machine then says, okay, I figured out what you're doing. One application of that is, for example, a friend of mine, does work in, telemedicine and application of medicine.

And they would ask doctors, if you had a patient came in with these kinds of symptoms, what tests would you do? What would you do with the results? And the idea would be to say, this is the normal of behavior of what doctors should do. The problem with it was when you ask doctors what they would do if a patient came in, they would say, if a patient had this, I'd order these tests.

I would do these things. If those tests show up this way, this is [00:06:00] how I would diagnose it. They would write a decision tree. Then they videotaped doctors without their knowledge when a patient actually presented and what the doctor actually did. was in many cases completely different from what they said they did.

Then, afterwards, they would ask the doctor, What did you do? And the doctor would describe what he did, or she did, based on what they said they were going to do. And the truth was, there were a whole bunch of things the doctors did that they didn't bother to write down. Like he says, First thing I do is I would check to see how the patient was feeling.

I put my hand on their wrist to check their pulse or respiration, see if their skin was climbing. They wouldn't write that down because that was instinctive. Okay. They would look to see if the patient was diaphoretic or sweating or uncomfortable or looked pale. They wouldn't write that down because it's a naturally distinctive thing.

So one of the problems you have is humans are really bad at knowing what we do. And why we do it. There's a much that we do that [00:07:00] is based on the, limbic system on, on, on stink that we have embedded. So when we try to describe what we do, a terrible job of it. So when you mentioned that the, a book came to mind that you may have heard of called the Checklist Manifesto.

And if you take a look at what private pilots as well as commercial pilots, I'm only a pilot. Private pilot, not a commercial, but you do everything with a checklist. Even if you've been flying 20, 000 hours, you always use a checklist. And the idea is, that it is a way to take care of a complex process to ensure that you don't get distracted.

So for example, when I was packing for this trip, one of the things I told my wife, I said, just leave the room for 20 minutes because I fly a lot. I don't start packing until about an hour before I have to leave for the plane. Even though I'm going to be on the road for four weeks. And as a result though, it is a very precise focus on it.

And if somebody comes in, Hey, do you know where this is? Or did you take this out or take the water, the dogs or whatever, I'll miss something. And of course you find out about it when you're on the road and for the [00:08:00] most part things you need to buy you can buy, but it's a little bit annoying when you go man, I could have had this stuff.

The other thing you mentioned there that I thought was interesting is the idea of the limbic system. You get down here into the lizard brain where you get this reflexive action on things where it's the thinking fast and slow by Daniel Kahneman. Exactly. Where Kahneman And you're going, Wait, I didn't even think about having to write that down.

And so we struggle sometimes. For example, one of the things about the checklist for pilots, you get in a plane, you're going through the checklist and you smell something funny, smell gas, you smell CO2, you smell that's not on the checklist. So the problem with writing a checklist is you can't include all the things that we do instinctively and say that or the plane sounds funny.

Okay. Or, and it's not on the checklist. It's not something you would normally check for, but it just, it sounds funny or. We're at a higher elevation, so it should sound different, at a higher elevation [00:09:00] than at a lower elevation because it consumes gas differently. These are kind of things that we know instinctively.

And one of the things my friend who works with artificial intelligence was doing was he was working on early generation of self driving cars. And one of the problems you have with a self driving car without all the decisions you have to make, a lot of them can be, written, but as you're driving down the road, you see something in the road.

And you instinctively say, that's a log, and you say, I need to avoid that. It's a small log. I can drive over it. It's a paper bag. I can drive over it. It's a piece of a tire part. I can drive over it. And you make an instinctive reaction to which things you need to come to a stop for, which things you need to avoid, and which things you need, you can just go over.

And that's because we have in our brain, a database [00:10:00] of every object we've ever encountered, mostly, and every orientation of that, and can make an instant reaction to that. But if you were to ask, I would ask you, what is a chair? And then try to describe that. Then you could have, most of us think of a flat surface, which he'll have probably four legs.

It could have maybe three, but often we'll have a back, maybe arm rests, et cetera, et cetera, designed primarily for humans to sit on. Is a bench, a chair, this is the whole, is it as a hotdog, a sandwich kind of thing. And one of the problems is tomato or fruit. and you have chairs, beanbag chairs, the chairs are designed to look like a baseball glove.

And so we have a functional analysis of what a chair is. And when we look in a room, we can say, oh, that's a chair. It's very difficult to teach a computer that kind of stuff. And so when we talk about machine learning and artificial intelligence, [00:11:00] it can approach human thought. Now, one of the problems is if you direct something, you are embedding whatever you're directing with whatever biases or prejudice you already have.

Whether they're conscious or unconscious. That's right. And it could even be on the training data. For example, if you have a hiring database and for 10 years you said, Hey, we know exactly who we hired. over the last 10 years. So let's go ahead and run that as a training set. And now we'll start hiring new people.

it might've turned out that 10 years ago, because of where you were in the company or the type of jobs, you might've been primarily white males that were applying for jobs. And now today you find out that females or people of color don't score as well. And it's not because you're trying to go ahead and discriminate.

But the database had a skewed input, and that's all the computer is going to know in a directed model. Now, you mentioned an undirected model. How does undirected the other problem, with directed models is this. There's a tyranny of what can be measured, right? [00:12:00] so when you're talking about a directed model or an undirected model, you're talking about data analytics.

That means it's something that can be actually measured. And what an example is you want to, you have a baseball team and you want the baseball team to win and you have two choices. One is, you can train the computer into what are the rules of baseball, what is the object of baseball, traditionally what teams won, and look at whether it's better to try to steal bases or not steal bases or bunting and things like that.

Or you can say, I'm going to feed into this computer every baseball game ever played. It's going to figure out what baseball is. It's going to figure out what the rules of baseball are, it's going to figure out what the object of baseball is, and it's going to tell me what is the things I should be doing if I want to win a baseball game.

Now, in the undirected one, where I just feed all the data in, I don't tell it what data points are important. I don't tell it, [00:13:00] it figures out what the rules are after watching a lot of games. And it comes back and says, I have figured it out. The number one thing you should do if you want to win a baseball game is wear blue socks.

Or chew tobacco and spit. Or chew tobacco and spit. Now what's really interesting about undirected AI is we, as students of baseball or whatever, will say, oh, that's stupid. And we will reject the results because it doesn't comport with our biases or our biases of prejudice. And it may turn out that yes, it's right, but we're not going to accept it because it's stupid and makes no sense to us.

And that's where the idea of AI being creative comes in because really AI, particularly generative AI, is just simply mathematical modeling where your tokens are often words. So something like a ChatGPT is just trying to come up with the most logical next word in a sequence. And if I say peanut, you might say peanut butter.

There are other things, but peanut butter, and then keep going, peanut butter. And [00:14:00] peanut butter and jelly, peanut butter and jelly sandwich. And if you say, do it again, you might get peanut brittle for Christmas gifts, or then you might get peanut allergies or, and eventually what's going to happen is that if you have a dive deeper and deeper into these lower probability associations, it comes up with something that you and I, as a human professional would have rejected long before we got to that, because it was just so counterintuitive.

But you look at it and go Wow, that's brilliant. It's created something. Really all it did is it just took the data we gave it and came up with non obvious relationships that we would have rejected out of hand. One, one of my favorite things is I have a game on my computer called Akinator, and it's a form of 20 questions.

And you think of a person or place and thing, and it asks you questions, try to figure it out. And what's interesting about it is that The way it asks questions seems completely illogical. so it'll say, is the person over 40 years old? Are they an American? Are they a real [00:15:00] person? Okay. All of that is great.

Then it'll ask some weird question. Whereas if you said okay, it's a person over the age of 50 that are associated with sports in New York, it's baseball, our mind will start from the broad and try to work our way narrow. It will ask a weird question, do they have red hair? And that's not the way we would narrow it down, but what it's doing is it is using its own algorithm to figure out what is the best question to narrow this to at least the fewest number of questions.

It's not the way a human thinks. So in one way, it mimics human thought. And in another way, it doesn't. Another example of the problem with AI is that AI has a very bad time dealing with outlying data. So an example would be a woman. goes for blood test and the blood test comes back with a high PSA, prostate specific antigen.

Okay. Now, if a doctor got that result on a genetic [00:16:00] woman, all right, what would the doctor assume? That there is a crossing of the samples. Yeah. Somebody screwed up. Somebody screwed up. He would reject the test, order the test again. The problem with AI is trying to fit the data into the model.

So the AI program will come up with the one obscure disease whereby women can produce prostate specific antigen, okay? Whereas we, because we have experience, say the most likely scenario is it's bad data. Alright, so that's another problem with AI. So you're dealing with the garbage in problem. The data that we're using is biased.

Many times in ways we don't know. It is incomplete many times in ways we don't know. It is, biased by the fact that it has been collected and measured, which means it didn't collect all the stuff we didn't collect, right? It may be old, it may be contradictory, okay? It may be inconsistent, it may be measured in different [00:17:00] ways.

So we've got A bunch of bad data. And it doesn't know what it doesn't know. As humans, we know, hey, I need to go get some more data. I go look it up and then I fill in the blank. But the AI model might not know that it doesn't know something. That's right. It is confidently stupid. Time for a hallucination.

The other problem is from a legal standpoint, the training materials, the things we are training in on, are owned by somebody. They are copyrighted. They may not be. They may be. And so now, if I were to tell somebody, a human, paint me a painting of Times Square in the style of Claude Monet, and they make me a painting, it's pretty clear that the artist that I hired owns that painting.

Maybe I own it as a work for hire or whatever. But when I added the words in the style of, did that make the work a derivative work of Claude Monet's? Does the estate [00:18:00] of Claude Monet own any rights to this painting created 100 years later, 150 years later, based upon this person's impression of an impressionist?

And by and large, the answer was no, you could create a work that was in the style of, as long as it wasn't too close to, right? In the area of AI, we don't know the answer to that. If I say to a program, write me 10 jokes as if they were written by Sarah Silverman, and I pick her because she has lawsuits against this, it will do that.

Does Sarah Silverman, a living comedian. have any rights to those because I think they are virtually indistinguishable from what she created. And are those rights superior then to the estate of somebody who is deceased? one of the things is that you, a copyright is usually good for a limited period of time.

And As we're finding, as Disney's finding out. As Disney is [00:19:00] finding out, the Mickey Mouse rule, right? As it learned only recently. The, old rule for copyrights was that the copyright would be, last for as long as the copyright to Mickey Mouse It was like, it used to be life plus 50. Then he moved to 75.

And in fact, if you remember who the congressman was who pushed for the McGee, Ms. Sonny Bono. Sonny Bono, the late Sonny Bono. And that was a big push to be able to extend that. it turns out that Mickey Mouse is still copyrighted, but Steamboat Willie is not. And so some people have come up with some pretty weird cartoons of Steamboat Willie with the chainsaw, et cetera, figuring that there's no restrictions on that.

So 75 years after you're gone, it's fair game. there, there are all kinds of ways to extend and, and do that. But by and large, a copyright is intended to be a. A monopoly, a license that you have for a limited period of time. Within a copyright, you have certain uses that you're allowed to make and certain uses that you're not allowed to make with a copyrighted work.

[00:20:00] And can I use a copyrighted work to train an AI program? Forget what the output of the AI program is just to do training of the AI program. Can I have a computer read every book ever written? can I go ahead and buy the book so I have actually compensated the author in whatever mechanism at whatever price the author has chosen, and then train on it?

Because at that point in time, I would think that if I'm a kid and I'm doing a book report and I have to read, Light in August by William Faulkner. If I write a book about what happened in that particular essay on that, I have to draw from that. The short answer is no, because when you buy a book, what you're buying is ink on dead trees.

Okay. You're buying a license to read the book and quote from it in certain ways, but you don't own the intellectual property in the book any more than I own the music on a CD. And that's the other problem is we've got these rights that are embedded within physical media and then rights that are, [00:21:00] embedded, that are separate from physical media.

And the Supreme Court had a case a number of years ago about somebody who was buying cheap textbooks in the Far East, in Asia, and selling them in the U. S. market. And because it was physically buying physical textbooks, even though, so there was a two market system. These books would cost 200 in the States, and you could buy them for 20 in Indonesia because it was a different market.

But there was a limitation on resale and the guy was buying them in Indonesia and selling in the United States. And he was sued and the Supreme court said, I don't know what's called the first sale doctrine because he'd bought them there. He could resell them because he was, what he was reselling is the physical book with the, and they'd already been paid for it.

So that's your argument. What he didn't own was the intellectual property, so he couldn't print his own book. And that makes [00:22:00] sense. So how about medication? So you have all the same issues there, but now you add FDA regulation and the like to it. So you get into, could I sell a product here and name it? now you get into trademark issues as well.

Or the Canadian pharmacy where they say you can go ahead and buy your, whatever it is that you're looking for, which may be lead us into the next one, the AI Gen porn, but the idea you can go ahead and get. Little blue pills for less because they come from across the border, even if they're seen something else, which is called the monkey selfie Okay, the monkey.

So monkey self. Okay. I've not heard of that. So it's actually not a monkey, but, but I think it's a macaw, but there was a case involving a macaw. They gave him a, a camera, a digital camera, and he took a bunch of pictures. And he, one of them he took was a, a selfie. And the question was, somebody posted that on the web, and the question was, who owns the copyright to that picture?

It has to be a human. no. Or a company. [00:23:00] The first thing is the creator is the one who owns the copyright. And that's a real problem in the area of digital porn, because if a boyfriend takes a naked picture of the girlfriend and then posts it online, she cannot get that removed under copyright law, because the holder of the copyright is the person who took the picture, not the person who's depicted in the picture.

Advice, by the way, to our listeners. it, is a major problem about, revenge porn and the like. and so some of the remedies that you have under copyright law may not be available, but there has to be an author. And in the, Monkey selfie case, there was no author, no human author.

So you cannot be copyright. So was not able to be copyrighted. Okay? Bring it to ai. Now you go to ai. Now, if I ask AI to generate something for me, write a 1000 word article on the subject, on this subject, number [00:24:00] one, is it a derivative work of all the stuff that it researched? And number two, is it even possible to hold the copyright on that?

Because it's interesting because you think about it. And what I have read out is questions like on the copyright holding that there's a couple of ways. We talked earlier about life. Plus 75 for the company. It's 125 years. And so if you're going to have a company, you start, you're going to live 50 years longer than your company.

You're probably better off with the first, otherwise you get the second one. But ultimately copyright laws, as you'd said, it's a limited monopoly for the creator of something or somebody who buys the rights to something. But it does bring up an interesting question in the world of AI as we get more and more things that are coming out of the AI genesis is that Is it the owner of the data that it trained on, which appears to be the lawsuits that are going against the AI companies?

Is it the owner of the model builder, the person who spent a bazillion dollars [00:25:00] in training it? Is it that's right? Is it the person who actually wrote the code who spent years writing the code? And then. He had the code, the training database, and then the actual run of the training, or is it the person who actually created the prompt?

Write me a romance novel in which the, the, this is what happens, Somebody on a beach, they fall in love, and he eventually breaks her heart. That's the whole prompt. And I said, write me a romance novel based on that. I'm doing very little work. All the other things you're talking about is doing the heavy lifting.

If at the end of that, I say, you know what? No, here's what I want to have happen. In Act 1, Scene 1, they meet here. This is what happens. This is what he says to her, blah, blah, blah, blah. And now it generates. And is it fill in the blanks so that you go from an outline? And now I say, okay, yeah, do that.

But this time, make sure that she is more of a dynamic figure. He's more of this. And then I'm crafting something. Now there's a lot more [00:26:00] input by me as the prompt writer. does that, Take it over the threshold. So now it's my work rather than the AI's work. Or what about, let's continue it one step further.

Okay. AI produces a story and now load it into my Microsoft Word and I start editing it around and I change this and I move that around. And now I've made some alterations based upon me taking my idea prompt, running it through an engine, which. gives me some working time. It's now I have a piece of, I mixed up the dough and I rolled it out.

Now I'm coming in with a cookie cutter, cutting out only what I want and throwing away the rest. At some point, it becomes your work. It becomes my work again. And the answer is, we don't know when that is. And that goes to the fundamental core question is whether or not the output can be copyrighted at all.

And if so, what Who owns the copy? And that becomes important because one of the things you have, from legal and legal agreements is confidentiality and you say, who owns what within the relationship and when stuff is generated by AI, we don't know the answer to those questions. and [00:27:00] applying that in the security field, there are all kinds of cool things you'll be able to do with AI, one of which is you can say.

Based on all my documents submitted, write me a robust information security policy based on the law, the regulations, and what I am actually doing. So you can now have a tailored information security policy. Write me an incident response plan. Write me a dynamic incident response plan based on changing circumstances.

I'm in the middle of an incident, what should I do now? And then the question is who has liability if it's wrong? What if I follow the instructions? So we are literally 18 months into an AI journey that will take decades. That's interesting. Cause you talk about policies and things such as that. I was, I do some mentoring and try to help people with their careers.

And I was talking with a lady this past year up in the Washington DC area, who has made her career, wanted her career to be about writing policies. [00:28:00] And IT security policies and this, that, the other thing. And she had this pricing model where she thought she's going to make this much money. And then as Gen AI came out, I said, here, I've got this thing.

And all of a sudden her number is getting down lower and lower until finally like you need to find something else to do. Because although There is value if you want to say, I want to hire somebody like Mark has been around for 35, 40 years in this career, who understands it, who is an attorney, who's got this great training, who's got this great background.

That's a bespoke type of a creation that I might want to create and pay for. and here's the interesting thing. So we, if you ask a generative AI program to write me a comprehensive. Incident Response Program. The way it's going to do it is it's going to look at all the incident response programs it has available in the training module.

And it's going to say these are the common things, blah blah. What it doesn't know is which ones work and which ones don't. It doesn't know, wait a second, that particular phrase was a problem six years ago in this litigation. But eventually it's going to learn all that stuff. [00:29:00] It'll get smarter and it'll get better.

But what it will not do is it will not get more creative. It will be able to reproduce the things people had thought about in the past, but it will, it is much more, it's easy to say, write a painting in the style, make a painting in the style of, Claude Monet. It is difficult to say, write a, make a painting in a new style that nobody's ever thought about.

End. From that perspective, if you define a style, I like broad brush strokes. I prefer to have primary colors. I'm looking for something that would tend to be a watercolor. So it's going to diffuse. And all of a sudden, if you don't have very good artistic talent, Could you still become an artist if you're able, to go to something like a Dali and give it enough detail where it could interpret that and say, I've seen watercolor.

I've seen primary colors. I have seen Andy Warhol stuff. I have seen Monet and now this is yours [00:30:00] and you can call it. I think at some point it becomes unique enough. And now the problem is whether you can prevent other people from making artwork in your style or music. Exactly. And when you look at creative things such as artwork or music, et cetera, is compared to a chair, which is a functionality.

You can measure the functionality of the output of the chair. If you sit in it and you fall to the floor, it's probably not very functional. But is music functional? Is art functional? music and art both evolve. They serve a function. Okay. and it can be an emotional one. It can be an expressive one.

It could be a communicative function. There are lots of functions that it can conserve, but good music. I think everybody agrees on what's bad music and nobody agrees on what's good music. and the other thing is, when you see it though, I'm of the generation where my musical tastes got locked in, somewhere between April and November of 1976.

And therefore [00:31:00] anything written before that is, accepted by me. And anything in the style that was accepted back then is good music. And everything else is measured against that. And we all have that point where our musical tastes get locked in. It doesn't mean we can't appreciate other music, but it means it becomes measured against that.

What is difficult, as creative as AI can be, it is difficult for it to come up with something that is different from what it has been trained on. And that's because of the fact that it's at the end of the day and just a bunch of numbers. It's a mathematical relationship between what pairs or triplets or whatever, as it's seen in the training data.

And then it's going to simply say, Hey, I have seen this, but it's like George Carlin had a routine years ago. Remember things that you'd never hear. Hand me the piano. When they probably went downhill from there, but the idea was it was non sequiturs that we would laugh at the clairvoyant.

Some people will have to look that one up to see who used to say that, [00:32:00] but we might remember Steve Martin. And so as a result, you can use these tools to create. Things that you would expect to be close enough to the former art or the state of art, whatever the state of art happens to be, whether it's writing a book, playing music, et cetera, drawing artwork, but it's not going to go off on its own.

And that's one of the things that is human. Or it might, and in ways that are unexpected, And then the thing is that, a deterministic or a non deterministic process because at the end of the day, and that's always been the problem with random number generation is computers don't create random numbers.

They do pseudo random, but unless you have some sort of. Random source such as cosmic rays or radioactive decay or something like that. You can't get purely random. And that's been a crypto problem for years. You pointed out one of the big problems with, AI is oftentimes let's say you have a, an AI program that is assisting in telemedicine or [00:33:00] remote medicine, whatever else.

It's when it screws up and somebody dies, one, it says, whose fault is it? Okay? Second thing is, what you do typically now is you will check the code. and say, okay, what did the code say? Blah, blah, blah. There's no code to check in an AI program. There's nothing that you can go back and look at and check the code.

And as a result, it's difficult to determine why an AI program did what it did and whether it in fact screwed up. So from things, when we have AI embedded in a whole bunch of stuff, we're going to have issues about AI and product liability, AI and negligence law, AI and standards of care, and that's. Even putting aside the issues about training models and what it was trained on, even if it does what it's supposed to do, bad things will happen.

Quote

And the other question is about a standard of care. So people, [00:34:00] you're designing an AI program to develop a self driving car. How safe should it be? Should it be about as safe as the average driver? About as safe as the best driver? or perfect And then when you get to a point where you have to make a choice, do you preserve the driver's safety or a pedestrian?

Which means if you swerve, you go over a cliff. So this is the trolley problem. The trolley problem. And what's very interesting is these kinds of decisions about the trolley problem are actually embedded in technology. And I'll give you an example. In Europe, cars are designed with soft bumpers because there are lots of urban areas and small towns and small cities, and you're more likely to have a car to human accident.

And they are designed to reduce the impact on the human. If the human gets contact with the car in the [00:35:00] United States, you're more likely to have car to car accidents. And therefore the bumpers are more robust to protect the humans in the car from that kind of accident. But that's at the sacrifice of a human to car contact.

And we have to build that into the technology. And so we have to literally build in the trolley problem into the technology. That's interesting. I remember I. Years ago, I had a 73 Buick Electra and it had five mile an hour bumpers. There were these huge pistons and you just go ahead. And if you were to, you could drive it and take it into neutral at five miles an hour and the front of the back were in different times and it seemed that way.

And, oh my goodness, this thing was huge. it got about 10 miles of the gallon. I guess people didn't care back then. I bought it from my old college roommate for like 300 bucks. I drove it for a while and when it finally died on the New Jersey turnpike, just took the license plates and said, yeah, tow it, keep it.

that's the whole thing is, ethics. Ethics go to the training modules, they go to the underlying [00:36:00] data, they go to the decision models, they go to the hallucination issue. And When we talk about ethics in AI, one of the big problems is we don't, there are all these competing values that we have as humans.

And how do we train an AI program what to do that is ethical when we have competing values? on one hand, we want to reduce the illegal immigration. On the other hand, we want, people to be able to have, be able to, be their best people and not starve to death and stuff.

How do you value those two? On the one hand, we want to punish people who've committed crime. On the other hand, we're, we don't want to kill people. How do we compare those competing values? And a lot of them are the way how you phrase the value, if you phrase this as freedom or liberty or that kind of stuff, it's going to get one set of, responses.

Marker

So the best we can do [00:37:00] is hope to mimic. Human behavior. And again, people lie about what they do. They don't remember what they do. And when I was talking about, when I was talking about the, the self driving car is we are much less risk tolerant when we have to describe what we're doing. So if you had, if you designed a car that drove as well as most humans, the peak of the bell curve, that AI program will kill 40 to 50, 000 people a year in the United States, because that's how many people die.

And we're going to say, wait, we are having litigation right now over two people dying in Tesla crashes. All right. And so you have to be willing to accept that to the extent we want AI to replace what we do in human decisions, it will be wrong. What's the acceptable level of loss, which is now we're back to looking almost like they're cyber models that we say, when [00:38:00] can you get it wrong and by how much?

And if you allow a certain hemorrhaging of financial information, customer information, integrity, confidentiality, availability, but we say that's an acceptable loss because we can accept those risks because it's not a life threatening thing. And normally when things get really screwed up, when things go really badly, it's because a bunch of things have happened that we've never seen happen together before.

And it's a convergence of scenarios where one would think, can a computer, using that term broadly, assimilate all this data and then come up with the right answer. And the answer would probably be only if it's seen it before, because if it's absent from the training data, it's going to be sucking wind.

And then because these Gen AIs are like a seven year old, it'll always answer your question. I have no idea what the answer is, but it will give you an answer. It's going to go instead of the 0. 1, 0. 2 to the 0. 0001. The only thing it can find That could be a [00:39:00] non sequitur. It could be something out of science fiction.

It could have been anything. If you look at the CrowdStrike problem, five or six things had to go wrong for CrowdStrike to have been able to do that. I'm not sure an AI program would have been able to predict all of those things. The perfect storm, if you will. The other, there are some things within information security that I think AI is going to do a fantastic job of doing.

One of the things AI is really good at is looking for patterns in massive amounts of data. And when you think about things like intrusion detection, intrusion prevention, IDS, log monitoring, all that stuff, including behavioral analysis of malware. Awesome. Yes. Okay. It's going to do a great job, but remember that these attack systems are iterative.

Somebody is trying to defeat them. We're not against a static pattern. So people are using AI to say, Hey, I want you to write malware that will defeat an AI prevention [00:40:00] system. So there will always be this radar, detector, radar jammer. And it's the same company that sells all of them.

But I think in terms of the normal day to day stuff, that's an area where I think AI is going to be fantastic. The second area is in displaying information to humans. What do I need to know? And it won't know what you need to know, but it will know what it has displayed to you that you acted on last time.

And it'll say, based on what you were acting on last time, this is what you need to know. The third thing is, that I think AI will be really good at is, Pattern, anything that requires pattern analysis and a lot of what we do developing pro incident response plans based on real things rather than what we think, writing policies that are geared towards real things, data flows.

Where's [00:41:00] my data flowing, right? Or as we learned from the Israeli pager problem, looking deeply, at the supply chain issue. And so a lot of challenges coming up, of course, we're getting down to the last couple of minutes here in the episode. And as always, I like to have twice or three times as much time.

So we may do another one soon. I think we will. I think we will. But. If we take a look then, it's wrapping things up. So we talked about AI and the law. What is AI? We had talking about directed, where we go ahead and we tell it what to figure out. And we had undirected where we just said, have at it and see what you can figure out.

we discussed copyright and then things like AI generated products, who potentially could own that. Can you copyright it? Whether it's an AI device or a Macaw, who owns that output? And then ultimately, can you assert those claims on it? And then. Can you call it your own? You can call it your own potentially if you have enough input into there that one would say that was such a creative and complex prompt that nobody else could have possibly generated that.

[00:42:00] And who has liability when AI misbehaves. And when it goes off the rails. And then lastly, I would think is that as a bit of a cautionary tale is if you look at an AI tool, even if you were to be able to do a comprehensive code review of the AI itself, I suggest that's insufficient and it's off target because you could train it with data that says lie, cheat, and steal.

When the clock turns 2025, give the wrong answer, but during 2024, when everything is being trained, give the right answer. And all these things could be embedded in there. And unless you had a, Utterly comprehensive testing model that would try every possible permutation of inputs. You're going to be able to embed stuff in there.

This is the malicious compiler problem. you have code that looks good. You compile it, and then the compiler itself injects malicious software into the code. So the output code is evil. When you go back to [00:43:00] examine it, you decompile it. And it comes back good, and there's no way to know when an AI program is misbehaving.

And that is even before you even get into the whole hallucination problem. So a lot of challenges ahead. I think we're at the cusp of something fascinating in terms of AI gen AI. We're going to see, even news recently about, hey, let's go ahead. And there's an effort to go ahead and take some non profit companies profitable and just think of all those zeros involved in things such as that.

And, it's something where you shouldn't avoid. It is a recommendation I have to people go out and invest it and spend the 20 bucks a month and go ahead and buy a better model. Try to program your own. There's a website called thereisanaiforthat. com that lists 100, 000 different AI models of various qualities and permutations, but it's a kind of a google listing of all the AI models that have been identified for it because they're all going to operate a little bit differently and be careful before you do anything that [00:44:00] involves life safety.

At the end of the day, as a human, you should maybe trust but verify, and that would be important. Any last thoughts? at the end of the day, with life and safety, we, I feel more comfortable the human making a bad decision that I can understand than an AI making a good decision that I can't. Good words to end up with.

So anyway, thank you very much for everybody listening or watching out there. This is G Mark Hardy, your host for CISO Tradecraft. I've had on my show Mark Rash, Esquire, a dear friend of mine that I've known for a long, time, and I'm privileged to have him back on the show again. And if you liked our show, please go ahead and give us a thumbs up or five stars, whatever rating it is that are out there.

We're not really grade grubbing, are, but the idea is it helps us reach more people. But until next time, thank you very much for tuning in and stay safe out there. Stay safe.