*Barclay Damon Liv*e Presents *Cyber Sip™*

Season 3, Episode 2: “The Boom of Generative AI in 2024—Breaking Down New Security Trends,” With Siwei Lyu

Host: Kevin Szczepanski

Kevin: Welcome back, everyone, and welcome back to you, Professor Siwei Lyu.

Siwei: Thank you. Yes, thank you again.

Kevin: Oh, it's a pleasure. Back by popular demand. So let me tell you a little bit about Professor Lyu. He is the Empire Innovation Professor of the Department of Computer Science and Engineering at the State University of New York at Buffalo. His research includes digital media forensics, computer vision, and machine learning. And we're going to talk about all of that, or at least some of that today. Professor, we're going to focus on AI and generative AI. We'll talk about some digital media issues as well. But before we get started, I don't think we're that far along in the world of AI to assume that everyone knows what “generative AI” is. So I know that in language models, it's a model that predicts the next word in a sentence based on the previous word. But I'm guessing that's a pretty simplistic explanation. So what is generative AI?

Siwei: Well, generative AI is a new trend in artificial intelligence, and it was actually used. It was very it's relatively recent. And that the time we use generative AI is in contrast with a more traditional form of AI work, which I will call “analytical AI.” So in analytical AI, the target is to basically come up with an answer from an input, say, you know, somebody asks a question and it’ll say, if this image contained the human face or is a cat or it's a dog, then the AI model will tell you yes or no. So basically, given an assertive answer to certain questions. What generative AI, on the other hand, is making up stuff so you can tell the AI model now instead of answer a question, say if this sentence contains a question about a certain opinion, is this, like this comment is an Amazon review? Means this customer likes the product or not? On the other hand, generative AI will say write me a comment about this product and expressing my like… appreciation. I like this product. Write a piece of content like that. So generative AI is a new trend of AI of making things instead of just answering questions.

Kevin: So I appreciate that, thanks, Professor. And that makes sense to me. When I think of analytical AI, I'm thinking that our research models: Lexis, Westlaw, Casetext, where I can insert a prompt and there is this corpus of case law and using a Boolean search or a prompt, I can then receive from that model…

Siwei: Yes

Kevin: … the most relevant cases, but they're not…that model isn't creating anything that doesn't already exist. So let's talk about generative AI that creates things that don't exist. And my next question is what happened in last year, last summer? No one knew anything about ChatGPT. No one knew what generative AI was or is. And then in November 2022, OpenAI launches ChatGPT and now all of a sudden we all have to figure out how this is relevant to our lives and how we can incorporate it into our lives and businesses. I just… I know it didn't happen overnight. Can you just walk us through how it did happen? How was it that last fall we all became introduced to generative AI?

Siwei: That's a great question. I think all of this is actually an accumulation of many years of research in this area. The public only become aware of generative AI, starting, as you mentioned, November of last year. We probably just passed a bit over one-year mark of the development of the… I'll say, coming to the public awareness of generative AI. But this area of… branch of artificial intelligence has been under very active development for at least 10 years, not if not even more. So people…this is actually when we started the artificial intelligence, analytical, analytical AI and generative AI are two branches going in parallel and sort of borrowing area from each other and encouraging each other, promoting each other too. So these are the two basic mapping back to two basic activities of human intelligence where human brains are dealing with. You know, in the real world, we have sometimes have to deal with analytical stuff or how to make decisions on a daily basis. You know, should I turn left or right? You know, I just crossing when I'm driving. But we also need to come up. We also need to do a lot of things that are creative, requires, you know, original thinking. And those are generative, intelligent behaviors. So this is… ever since AI, artificial intelligence, becomes a research area, people start to get interested in both analytical and generative branches of it. So I think the earliest example of generative AI I can think of is probably from the early 2000s. When people are making AI models to recreate patches of images that look like real images. It was very primitive, but it was a like a baby step. But it's a revolutionary because back then people already… researchers have already started to think about how to use AI models to recreate what we experience in the real world. And what happened is, you know, that has most of the development of generative AI have been behind the scenes for…number one. The results are really not that well, you know, good to be presented. On the other hand, I think this also the limitation of computation resources and data. And many of the other technical factors. So is all of this development in this… like in many other fields, especially in the AI, where the development is like exponential. So before we reached certain threshold, it becomes this is always under the scene, you know, beneath, under the surface and there was last year on the one year last November when OpenAI finally opened up this ChatGPT that would be make it a public. A lot of people can play with it. Then we're suddenly aware that we have that capacity. But the precursor of ChatGPT, which is known as the GPT model I have tested and, you know, played with the system, I think as early as 2020 and they have a GPT version 1.0, 2.0 and ChatGPT is based on… 3.0. Sorry. So I have seen two previous versions of that and I can literally see the improvements of the generative quality across different versions.

Kevin: Yes.

Siwei: So, so there are a lot of development progress, allotments behind the scenes under the surface. And November, the November 2022 was the time that, you know, the technology is finally sophisticated and mature enough that the users… common user can start to appreciate it. All right.

Kevin: The moment that we all saw the tip of the iceberg—and I've use it as well. I've used it to create lists of questions more or less out of curiosity. I know some have rather infamously used it to conduct legal research, which has turned out to result in “hallucinations” that have gotten more than a couple of lawyers into some very serious trouble. We're going to talk about some of the dangers of generative AI in a moment, but first, can you just give us a sense… I you’re your focus, although you're very familiar with large language models. Your focus has been in in the digital media realm. Give us some examples. If we were looking at…you're predicting. So we're sitting here in January 2024 and you're looking out onto the horizon, what are some of the uses that you expect will be mainstreamed as opposed to uses that are confined to large corporations and maybe research institutions like UB…

Siwei: That's a wonderful question. And, you know, if I have a looking glass into the future, I will say, you know, definitely one of the major trends will be the personalization of the large language models. So, you know, right now we're at the mercy of whatever, OpenAI come up with their chatbot model trained on the whole corpus of Internet. But I think, you know, to make ChatGPT of large numbers models more useful for ordinary users and I would say particularly professionals, you know, like lawyers, corporate executive whatever have a special need of, you know, context. ChatGPT with you will become more tailored to that sort of quality. So, you know, like, you know, I want I want a ChatGPT like a personal assistant to me and be able to help me writing simple emails or template formatted documents in my writing. So then I think is going to you know, I will have such a GPT model that will learn from all of the materials I wrote in the past and then come up with something that really reads like what I wrote and that will make the model a lot more useful than the current level. And this is already happening. GPT OpenAI is actually opening up an AI where you know you can provide your own training to the stars and personalize or specialized the GPT model to come up with things that you know more aligned with your particular requirement and the right context. So I think that's definitely one trend. The other trend is you talk about, you know, interface between different modalities. So just coming up, come up with text is one thing. But we also want to understand, you know, interfaces between modalities because the social media, the Internet is by its very nature multimedia. So we have not only text, we also have images, we have audios, we have videos. So I think another big trend that I will see, I would predict is this interfacing between different media modalities from images, videos and audios to text, for instance, Chat GPT current GPT-4 model already have a visual module where you can upload an image it will say something. You can, it can answer your questions about the images. I'm assuming this will become very useful is especially for, you say, you know, for particular legal contacts there is a recording of a court case like two hour long, but we may not have time to actually see through the two-hours video so we can ask models like laugh and smile to say give me a summarize highlights of this video of two hours down to like 5 minutes. So I'll be able to read through this and also, you know, take out highlight moments from the video and compress that into short highlight video so I can watch that part. And also accompany that was short text description of all these highlights and these things can also go the other way. So you know we already have models like tools, like a stable diffusion, me journey that give me texts to create images. Or we can also have something like give me text and create audios and videos. So, so this will also actively develop in the research committee. So I'm predicting, you know, sometime within the year of 2024, we'll start to see some reliable systems that creating short clips of videos that will help us based on the text input to and we have a say, maybe somebody will say, give me a video that Kevin says this and that, you know, just put a text. I write I want you to say certain things. And the… it will create a video of you saying those.

Kevin: I see. So when you talk about—and correct me if I'm wrong—when you're talking about interfacing between different modalities, that's one example. So I have text, but I want to enliven that text with audio or video, you can create a video or an audio that in which I'm actually speaking that text. And I may also appear as the person speaking that text. And I don't necessarily…so I don't even have to record it. What do …you what would you need technically to make that happen? I know we were talking a little bit about something like this offline. I was telling you that I had recorded a piece of micro content for LinkedIn, but I was driving, I was paying attention to the road, but I was driving and we decided, you know what? We're not going to show a video of me driving and speaking at the same time. Right? And you said, Well, if you give me a video or a photo of you, we can combine those resources into a video in which you're not in the car speaking, you're in your office or in another place. Is that what you mean by the interface of different modalities?

Siwei: Yeah, that's what… that's one side, you know, from text, from some existing contents to all your visual content. So there's several levels of that. I'll say the simplest level is you give me a… if I have a soundtrack of you talking, you just don't have you your, your visual, your video of speaking those words. There are AI tools called lip synching. So, you know, the easiest one is I can take a video, say the one we are doing now and then I can use that. Yeah. To can be used to create essentially you know the same set of lips, you know video to match up with the input the audio so I can take your audio track, run through the algorithm generates the you know you your beeps as you were speaking those words and pace it that means your video like this one and then it will have the impression as you were actually there, talking, you know, speaking those words and that's the easiest one to do. Or, you know, slightly more sophisticated I can take of a single image of you and I can use that algorithm to create again, the [leap shapes] and then I can recreate your face movement and the movement just based on one image. Their quality will not be, as you know, we would not be as realistic or natural as if I have the video, original video. But you know, for casual viewers, or you just want to put it on as impression of your talking there and it's pretty decent. So that's called “talking head” generation. Or even more, I can have just a text of you speaking certain things I can have your voices as training data. Then I can make a model that's called a textable speech. I can start… with starts creating a voice and then align your what's with the video? Okay, So I can now start with start to do that and even more sophisticated, you give me an idea I can run something like a large language model, creates a text and then make your voice speaking those words and then make your video and align with those voices. So almost like you give me an idea, say “make Kevin say these things.” I can go through this whole pipeline to make a whole video. You actually speaking extensively, not just give a full description of this idea and the end result of you talking. Okay.

Kevin: All right. So I'm thinking is, if I have a colleague who's no longer with us, but I have audio of this colleague making presentations and training and I should say when I mean no longer with us, I'm using the euphemism this colleague has passed on, so is no longer with us. But he meant so much to us. And assuming we have all the consents, we want to create new presentations with this person giving them, I'm no longer here, but I… you can create a video in which I'm giving a new presentation based on content that we can generate from text and it can look like I'm actually delivering that presentation, right?

Siwei: Yes. Indeed, yes, that's one of the most advanced generative AI tools can provide us.

Kevin: Yeah, but and we should I guess we should get to this now. But I suppose if in the wrong hands someone …there's lots of video out there of President Biden and lots of audio. Someone who is not out to do good could take that video and audio and create a speech in which the president declares war on one of our adversaries and that at the very least would create a tremendous amount of confusion, if not a response from the enemy country. So if that's the case, and if that's where we're headed, Professor Lyu, what guardrails do you recommend for this digital technology or generative AI? How do we come up with a set of rules and guidelines to make sure that… we're all running a thousand miles an hour, learning this new technology—that we use it for good and not for ill?

Siwei: I think, yeah, that's a question a lot of people ask me. You know, it's really a pressing need to make sure that this technology is being used only for beneficial purposes. I think fundamentally what this is…by the way, this is also the main objectives of the Executive Order from the White House for I think in October is putting guardrails to generating technologies because a lot of people are becoming aware of the potential negative uses of this technology, as you mentioned, spreading disinformation, using famous faces. And there are already real-world incidents of that. You know, I recently just got called on by a reporter from Reuters, and they identified fake videos of President Biden not talking about starting a war, but really talking about, you know, giving every American, you know, $5,000 for nothing, you know, just…

Kevin: Right.

Siwei: …is disinformation. It's not a… it's a fake video. And the with his voice, his face and looks real. So I think, you know, fundamentally, first of all, everybody, you know, making this kind of content need to become responsible for declaring the nature of the content we’re spreading. For instance, if I'm contracted to make a video of you saying certain things. I wouldn't just, you know, spread of it out online and telling everybody, this is what Kevin said. That will be the improper use of this technology. Why? And finally, I think what should always be done. There's always come with a disclaimer, with a declaration that this content is created with the consent from the subject using AI algorithms. So it's almost like, you know, we watch a movie… in the movie, you know, somewhere in the movie you will say, this is based on a real story. It would never say this is a real story, right? Of course, people come to a movie knowing that this is not the real event. But we should always tell them that, you know, something has been changed from the original sequence of events in the move. The same thing happened for the generative AI contents. We need to declare that this is not the real event, real video, but it's created with the help of AI algorithms. And at the same time, that’s from the responsibility part…from the technical part, the models we use it also takes some matters. For instance, the Executive Order the White House talking about putting watermarks into the syntactic contents coming out of the generative AI use. And also, you know, we can also put in certain fingerprints, extract fingerprints from the generative contents and store it somewhere. So future on when the synthetic content is being distributed we have a way to verify: yes this is coming from this model, from this tool. And also we should also develop detection technologies. That's my research. So that's in a short period of time. How could we make sure this content… if nobody declare it. Because, you know, not… we are not expecting everybody to become responsible, especially or talk about, you know, foreign countries or, you know, criminals. They're using that intending to use this irresponsibly. So in that case, we need to have a way to be able to tell, yes, this come from an AI, generative AI tolls, or no, this is actually a real video recording.

Kevin: I was going to ask you about that, Professor. Is there a way for you or your colleagues to look at a video and determine whether it's genuine or created by artificial intelligence?

Siwei: Yes, there is. There are ways to do that. We have technical detection algorithms, a bunch of them, so we can run through those detection algorithms of AI inputs, all the visual media and based on their signal …signatures, we can tell there are differences because again, you know, something come from a particular generative AI tools will carry certain type of features, certain type of attributes that we can use. It's just that those attributes may not be directly visible to human eyes or audible to human ears. So these algorithms are like x-rays. They can show those hidden features. We also developed along the years because I've seen a lot of this synthetic contents. I also do watch certain, you know, you can say I have trained eyes to look for certain things or to listen to certain features to see that. And that's really, you know, these things are significant fractions of my work is to help journalists and investigators to identify and exposed synthetic contents and provide evidence for that. Yes, we're working on that.

Kevin: Can you tell, for example, because you mentioned journalism. I wonder if you can tell, for example, in watching a video that's by and large going to be edited for a television broadcast. Is it possible to watch that video and using one of those algorithms you mention, determine whether the video has been edited, where it's been edited? So if you're looking at sometimes the edits are obvious, but can you tell whether there's been a subtle edit to a video so that people watching can know whether what they're seeing is actually a fair representation of what actually happened?

Siwei: Well, in principle we can. That's what this whole research area called “media forensics” is for. So were developing algorithms to help people to identify those subtle signs of editing or AI generation? But in practice, you know, it's a very challenging problem, and especially when the change are small and not very significant, the algorithm needs to be very accurate about where the video was actually edited or manipulated. So most of the detection algorithms are statistical in nature, so they give you an answer, but not deterministic answer. So it's not. …So the algorithm is used to help us narrow down the scope. Look for roughly this area or, you know, this one with sufficient doubt about is original. Then human expert coming in. Look for those evidence and come up with a more concrete conclusion about the nature of the video. So, yes, I mean, it's that is a combination. It's a collaboration between algorithms and human examiners.

Kevin: Right.

Siwei: But, you know, a short answer to a question, there are ways. But, you know, it's we're not there yet. That's why we're working on research.

Kevin: … and that's what people like you are working on. And I hope we do get there and I hope there's an interest in it, because I think you play an important role in helping us judge the value and the content of the information we get. We're running short on time. But anyway, in the time we have left, Professor, I wanted to ask you, you know, since ChatGPT came to market in November 2022, there have been a lot of well-known organizations rushing their competing AI products to market, and some of them are open, but some of them are closed in the way that you mentioned earlier that an organization may have its own AI, its own generative AI system that helps it perform its unique functions. So, I mean, there are there are upwards of a dozen. I think we may be up to almost 20 AI services now. You can get them from Google, Microsoft, IBM, Amazon and everyone listening may have a different type of business. So I'm an attorney in a law firm. There are health care providers, there are manufacturers. If someone sitting in a weekly meeting now, Professor, and they’re saying, boy, you know, I see all these AI products out there. Everyone else is looking at this. We should think about whether AI products can help us perform our business faster, more efficiently. How should an organization go about thinking about that and selecting a vendor? I take it…you don't just go out and say, oh, I just got this message from Microsoft. I'm going to order their product and we'll be off and running. But I think a lot of organizations just don't know where to start. And I know that many law firms are working on this issue now and other organizations, too. So, Professor, what advice do you have for us if we know generative AI is out there, we want to explore how it can help us perform various parts of what we do. Where do we start to think about whether generative AI can help us, what vendor we should use, and what guardrails we should have in place?

Siwei: Well, I think that's a again, very good question. I think to start with generally, AI you know, even though it's a very powerful is still a tool. So if the intention of using that to start with the actual practical uses so it needs to be grounded in real needs of the organization. So I think the first step is actually go back and look through the work, the whole workflow of the organization and look for stops where there are repeated formatted, and you know, requires certain repetitive editing, writing. Those are the places where GPT issues, especially like things like GPT, will be mostly helpful because the GPT model basically learned from a huge corpus of text. You know, this is called the training site and that training site is contains all the information it can use to create a text, or article. And if the more formatted the end result needs is the easier the model will come up with something that is readable and natural. So I think, you know, this is number one I will suggest is looking for the needs… where you actually need it, right, Because a lot of places is you don't need too much of variations. Most of time it’s just following certain formats and having somebody have to type in everything and write everything will become less efficient. That's where the generative AI tools, especially large language models will be most helpful. Now comparing different models, I think there are a couple of different factors you can think about. But you know, to me most of the…AI providers large language model becomes very expensive. So you actually do not have a lot of, you know, vendors who will have the capacity of running from scratch to a large language model and to provide it as a service. So between the major players, I would say the difference is actually quite subtle because it's really about the volume of data you use. And the general …with the architecture, they’re using are all sophisticated, complicated enough, complex enough that I think the difference are not significant. But the difference lies in how much they can be flexible towards your organization’s needs. Like can they open up and adapt quickly to a particular corpus that’s of your interest? Right. And number two is response time. You know, how well, you know, they can get changed along the way. So you put in… like in ChatGPT, you say I don't like this, change this style just a little bit. How fast a how many… in terms of number of iterations you need to make the model aware that the style needs to be changed to satisfy your requirement. I think number three is as you saw related with your last point is the safeguard is about “hallucination.” You know, when you give them the space of creation… making up things, how much of the stuff that will come up with that does not comply with, you know, set up rules or laws? Those are very important because the… I will say, if the model is more conservative and, you know, on one side it's less come up with some new things, but on the other side, it also reduced the need and effort from human to actually look at what it’s creating from the model and correct what is wrong. So I think this is a few factors you can you can throw in to decide on this. The best way to do that is have a specific set of tasks, one in parallel, see, you know, compare the results and the, you know, come up with data and then do… what you can do a better comparison there.

Kevin: You want to be able to compare each service provider with the common text just so you can compare apples to apples.

Siwei: Exactly. Yeah. Exactly.

Kevin: Well, Professor Lyu, thank you so much. I know you've got to run, but I think that generative AI is going to be a game changer for all of our listeners and our businesses across the country. But there are potential risks. And the more we understand about it, the more we understand how to use it responsibly and how to select the vendors that provide it, I think the more successful we're going to be. And the organizations that figure that out first are unquestionably going to have a leg up on their competitors.

Siwei: Absolutely. Yes.

Kevin: All right, Professor Lyu from the State University of New York at Buffalo, thank you so much for coming back. I really enjoyed talking with you about generative AI, and I hope you'll come back to talk to us again sometime soon.

Siwei: Thank you so much for having me. Yes, it’s my pleasure.

Kevin: Thank you. And thanks to all of you will be back soon with another episode.

Kevin: The *Cyber Sip* podcast is available on barclaydamon.com, YouTube, LinkedIn, Apple Podcasts, Spotify, and Google Podcasts. Like, follow, share, and continue to listen.

*This material is for informational purposes only and does not constitute legal advice or legal opinion. No attorney-client relationship has been established or implied.*

Barclay Damon Live*podcast transcripts and captions are automatically generated through artificial intelligence, and the texts may not have been thoroughly reviewed. The authoritative record of*Barclay Damon Live*programming is the audio file.*

*Thanks for listening.*