Barclay Damon Live Presents *Cyber Sip*™
Episode 37: "Exploring the New Frontier of AI—Everything You Need to Know," With Siwei Lyu

Speakers: **Kevin Szczepanski**, Barclay Damon, and **Siwei Lyu,** University at Buffalo

[Kevin Szczepanski]: Hey, everyone, this is a *Barclay Damon Live* broadcast of the *Cyber Sip*. Practical talk about cybersecurity. I'm your host, Kevin Szczepanski. Let's talk.

[Kevin]: Hey everyone, Siwei Lyu is SUNY Empire Innovation Professor at the State University of New York at Buffalo, where he serves in the Department of Computer Science and Engineering. He is also the director of the UB Media Forensic Lab and the founding co-director of the Center for Information Integrity. And Professor Lyu joins us now on Cyber Sip. Welcome.

[Siwei Lyu]: Thank you so much for having me on.

[Kevin]: Thank you for being here. I am a fan of your interviews and lectures on YouTube and I commend you and, the audience, if you have not seen them, tune in. There's some really great, concise and easily understandable presentations on deepfakes, which we are not going to talk about today, but we will in an upcoming episode. So thanks, Professor.

[Siwei]: Thank you very much.

[Kevin]: So with your permission, I'd like to talk about artificial intelligence or AI. Today, there is so much talk, sometimes frighteningly concerning talk on social media and cable news about AI. So I thought we'd start with the basics. I am not any expert. I know there is natural or human intelligence on one hand, an artificial or digital intelligence on the other. But let's start with the basic definition. What is AI?

[Siwei]: Okay. AI stands for "artificial intelligence," as you know, and is basically, you know, involves in designing algorithms and that drives computers can behave and do things the same way as a human will do. So it's a "computational simulation of human intelligence."

[Kevin]: So on that point, then, Professor, I did want to ask you, how does it work? At a high level, we don't have an audience of computer scientists, but can you just walk us through at a high level how AI functions? How do the algorithms facilitate this artificial intelligence?

[Siwei]: Sure. Well, if we thought about AI, the history is actually quite long. The first time this term AI showed up in... to the public was actually in the early 1950s. And AI has a long history and several generations of methodologies. So I'm talking about the most recent one, which is, you know, to be more specific, a lot of people heard of a name called "machine learning." So it's essentially, you know, equipped a computer and an algorithm with abilities based on this idea of learning. So how do we... how do humans learn some skills? You know, we do this by practice, like people say, you know, the way you get to Carnegie Hall is "practice, practice and practice," right? The same way is used for the machine learning algorithms. So what are we trying to do? Just taking a simple example. Let's say do image recognition. I want to recognize this is a picture of a dog and this picture of a cat. Now used to be, you know, to make the computer to do this, we need to tell the computer,

you know, cat is a furry, you know, with a tipped ear and that kind of fuzzy face. Dogs have longer, you know, snouts and, you know, other features. These are not computer learning. These are human understanding and transfer that knowledge for computer. And that is actually, you know, at a certain point of time, that's all what AI does. But it's less effective and we have to understand, we have to describe it in a way that computer can understand first. And that usually took a lot of time. Now, the new scheme of this machine learning is instead of just doing that, it's just like we teach our babies to recognize what is a cat, what is a dog. We point to an animal and say this is a cat and you know she understand this is a cat. And then we point to another picture and say, this is a dog and she understands this is dogs. Now she will figure out for the baby, her brain is very versatile. She will figure out what's the fundamental differences between a cat and dog. So in the future, when she sees something, she'll be able to make that connection to say, oh, this is a dog. The same thing happens for the computer, for the machine learning algorithms, which so many, many, number, many, many images of dogs and cats. And we tell the computer, we gave the computer feedback, this is a cat and this is dog. Every time the computer makes a prediction, we evaluate and our predictions. Say the computer, see a picture, say this is a cat we say no no no, this is wrong, this is a dog. So the computer will learn from the failures and build that knowledge internally over time. And eventually the computer can do a really good job. The algorithm went through a really good job of differentiating images of cats and dogs. That's how essentially, you know, algorithms learn.

[Kevin]: But I imagine that unlike human beings, it may take months or years or longer to learn to make those distinctions. The algorithm teaches the computer to do it much more quickly, if not immediately, right?

[Siwei]: Absolutely. Absolutely. This is because we now...one reason I should add that we are seeing this boom of artificial intelligence and machine learning is because we have unprecedented computation powers. So that's why, you know, even though the word "artificial intelligence" showed up in the 1950s, we didn't see a lot of the applications of AI back then because, you know, the computers are simply not powerful. We do not have that much data and the computer cannot, you know, process that huge amount of data. Now, we have reached that point and the computer can basically learn, you know, instead of actually reporting a picture to a baby, and that takes some time. You know, even those this to us is pretty quick. But the computers can read the image in a matter of milliseconds and then understand that. So just think about it. Condense the whole learning process of ...usually will take, you know, a list, you know, I'll say couple of days working with a baby. The computer probably can do it in just a matter of three hours. So that's why the power is accelerated. The learning process is accelerated, and that's where the power of this algorithm lies.

[Kevin]: Yeah. Now, many of us may have learned about the potential for AI only in the last few months. I think it was November 2022 when we all learned about a ChatGPT and the LLM. Can you walk us through the how the large language models work? How does AI work in that setting?

[Siwei]: Well, the large language model work almost in a similar way as the simple example I gave for image classification. What the AI, the large language model rely on is, number one, a huge amount of data. So in the case of ChatGPT, it basically for use all of the online documents, digitalized document, stopping at I think later 2020. So that's a huge amount of data you control. Number two, it's used immense computation power. So there are, you know, just gigantic computing our competition servers. There are big computer supercomputers and running for, you know, months to actually chew on all those data. And what it does is actually figuring out a prediction job as simply as what I just described, you know, I teach my daughter to recognize images of cats and dogs. ChatGPT uses similar ideas. So what it does, essentially predicting based on what I have... So you gave ChatGPT, for instance, a sentence, but it don't give the whole sentence. You give the first word. You asked, ChatGPT, and say give me the prediction or second word and the model will say the second word should be like this. And then we'll give a few options and think one of them will be most likely. And then you, as the human operator, tell them—now, we don't do this manually, but of course we have the ground truth. We have all the text. We can tell them the correct answer for the next word is this. The machine will take that. If it if his prediction is correct, it will take if you learn that. If it's wrong, most likely it will learn from ots failures in

as in the previous case. So that's how to build up its capacity slowly. But we're just talk about a massive scale of learning happening at the same time at a digital level, you know just as an amazing speed and so so that's you know if we if we take analogy of human learning process does based almost exactly the same it just you know expand it out to a different magnitude. Yeah.

[Kevin]: Now you have been...I see from your CV... you have been a professor in this field for better part of two decades now. So you're an insider. That being said, when you first saw this technology and how it works, how well it works, what did you think? Were you impressed by it? Is it... I'm sure it takes a lot to impress someone on the inside who is an expert in these areas. But what did you think when you saw it?

[Siwei]: Well, I. I think I have mixed feelings. To that extent. Well, as someone working in this area, you know, my work actually covered a lot of this large-scale models myself. You know, one thing I feel happy about seeing the results because, you know, finally there is a model that can do the job that will... most people will understand, you know, it's working well. And we have been waiting for that time for quite a long period of time. When I started my graduate study in artificial intelligence and machine learning, that was about 20 more than two decades ago. Back then the algorithm is so limited that, you know, we can only work on toy examples and, and anything if we want to work with real world examples, it's super hard. And that's why it's very, very difficult to convince people that these are things actually work. Now ChatGPT changed the whole landscape and impression of that. So I think, you know, for this area, for this research field it's certainly a great thing that is, you know, to help us to improve our overall image. So that's one thing. I also feel like, you know, this is a... kind of like not as surprised because I you know, as you probably heard about Moore's Law for computation, there is a similar words than Moore's Law for I think I'll respond to that. You know, with the increase in computation power and the scale of computation. Right. And all the large data I kind of already perceive at a certain point of time, we're going to see something like ChatGPT, know journey for images and so on. And so I think, you know, it's kind of like I'm less surprised and I'll say, you know, a typical user, first time interact with this. But on the other hand, I also have concerns now, you know, there is kind of like a face shift, you know, from people completely, you know, overlooks that you might put in AI know suddenly people realize, yeah, AI can be so powerful and there will be concerns there. So, you know, the concerns are not just, you know, starting from last year when ChatGPT showed up, although it was long running, you know in the academic circle people already talk about this just like people don't take, you know, other people to not pay a lot of attention. Think about this like a sci fi fiction that, you know, we're not there yet. And this future just happened very quickly. So suddenly, everything we talk about yesterday as a science fiction, today it becomes a reality. So those are two things I think I would like a researcher myself working this area. I start to ask a lot of questions because most of the time myself and my students are busy working on the technical details of artificial intelligence. And this is social impact that is a new phenomenon that we have to take, take it very seriously. So I think that's the kind of feeling I had at the time.

[Kevin]: So I know we're going to talk about this more in a separate episode, but since you raised it, I want to ask you about the open letter published by the Future of Life Institute. So a score or many scores of experts in the field published a letter in which they note that AI systems with "human competitive intelligence," as they call it, can "pose profound risks as shown by extensive research." And among those risks, at least discussed, are flooding our information channels with propaganda, automating away jobs that are currently done by human beings, and even the risk that we might lose control of our civilization. Now, again, I know we're going to talk more about this in a separate episode, but I read that letter and I've followed some of the press coverage and I thought, this talk of an existential threat seems fantastic. Just seems too, too bad to be true. But is it? And how many years away do you think we are, Siwei, at this point where AI poses an existential threat? Are we 10 years away from that? 50 years? 100 years? What's the truth as you see it?

[Siwei]: Well, I... I'm fully aware of the open letter you mentioned. I look at the situation a little bit differently. I think, you know, whatever prediction I'm making now, say five years, 10 years down the road, the letter, I'm going to dominate the world. You know, take the word "humanity." That was based on the assumption

that we're not doing anything. You know, we just sit here and see the technology walk by, is following its own course and doing that. I'm a strong believer that human... I mean, AI is powerful. The human brain is even more powerful. You know, we as a human, as a species, we survived, you know, 40 million years of 30 million years of history. Many, many disastrous, more disastrous situation happened. And we're still here. And one thing I'd learn is, you know, human are very versatile. We'll find a way to handle the problem. So, you know, I'm on one side talking about, you know, controlling, limiting the AI technology, I feel like, you know, we're even if we want to do that, it is probably infeasible and too late. Too late in the sense that the genie is out of the bottle. You know, there's no way we can put it back. I mean, maybe we can control, say, you, the US government and you know, all the Western democratic governments can refrain from, you know, using, misuse or abuse of AI technology. But we have no control over of others. I mean, only good people got limited in that set. Right. The other thing is yeah, technology are not I do not treat AI as evil, or good or evil because the technology itself is very neutral. You know, we shouldn't throw the bath water with the baby at the same time. Right. AI can do a lot of useful things for us. We may not notice, AI is already functioning, already helping us. Like every time you get a cell phone. Right.

[Kevin]: Give us some examples of that.

[Siwei]: Yeah. I mean, you pick up a cell phone. I mean, just it becomes so, so easy for us... when we take a picture, why there's a little box showing up, you know, telling you this is a human face and automatically focus your camera to adapt to that person's face. That's AI algorithm working behind on the chip of your cell phone in the camera helping you to locate where it faces are. And also, you know, you pick up a phone this time, you know, like you call your credit card company, you call your bank. You sure sometimes, you know, it's your normal times. The customer service phone calls are like, you know, almost like real humans. But they are not. These are systems developed in the 1990s where, you know, automatically recognize your voice and then generates voices for to answer your questions. So AI system is already there helping us it just like there are not like fanfare as much ChatGPT. I think the third thing is instead of focusing our energy on limiting the AI technology and, you know, putting emphasis on the potential danger, the o risk it creates, I think is probably more constructive to say how do we, you know, mitigate those potential risks? That's what I, you know, my earlier point that the prediction.. whatever prediction we have here is based on the fact that we do nothing. But we have a lot of things, a lot of things we can do. For instance, you know, we can encourage the right now, you know, AI developing in the current way is because that's where.

[Kevin]: That's where.

[Siwei]: ... Right. But can we make a financial model to make the companies, you know, aware of the social impacts to make sure their algorithm not doing bad things. Almost like I mean every time I think about this problem, I'm thinking about accountability. I'm thinking about liability, right? Like a car manufacturer. They have to pass all those crash tests. Otherwise there are costs. When I'll be on the market. Can we do something like that? Because you know, that way, if they do not pass this kind of tests, they cannot make any profit. Their product would not be even on the market. Maybe we should implement some measures like that to, you know, a guide to the companies behind the AI tools to comply with these regulations and to you know, I will say limit to the negative use of AI, but you encourage the good use of that. I mean there are so many good use of them. So that's the way I see it.

[Kevin]: I like the concept of it being a neutral technology too. So we're getting close to the end of our allotted time. I want to ask two quick questions before we go. First one is from my friend in podcast-land, Justin Daniels, who hosts the podcast with his wife, Jody Daniels, She Said Privacy/He Said Security. Maybe She Said Security/He Said Privacy. Either way, they cover them both. But anyway, Justin Daniels is kind enough to send me in a question. And the question is, you know, we're talking about ChatGPT and literally pulling in millions and millions of pieces of data, some of it personal data. How can we create a large dataset for AI and at the same time be sensitive to protecting the privacy of individuals who may be supplying some of that data?

[Siwei]: I will say we don't know how to do that at this moment. That is actually a very actively researched topic these days. This notion of privacy in big datasets are simply not... I wouldn't say none exists, but not being, you know, emphasized when we view it all the systems and... but that is a real concern. We talk about ChatGPT, it have some informations. I mean tax data is rather easier you can just red it out, all some red X it out, but that's all some of the sensitive information. But we talked about images is more concerning because, you know, somebody use my imagery to train a generator model without my consensus and later on they create images of me... of lookalikes of me. That is a concern. And also we have we seen recent cases of artists whose artwork was use in training of generator models. There be a huge argument about the you know, who right. You also you know the IP, were the copyrights belong to whom? So this is a huge open research area. We don't have any good solutions at this moment. But I think again, in the next couple of years, you know, there will be some solutions coming out. I'm pretty confident about that right.

[Kevin]: One other question before we go, Siwei. And on that note, if you had to make a prediction, what is this landscape going to look like over the next five years, 10 years or 50 years? What things should we be looking for? Are there going to be advances in economies of scale, performing jobs, jobs currently done by human beings are going to be done by computers. Will we see medical advances? What do we look for in that period of time?

[Siwei]: Well, I think I make an analogy of the wild, wild West now for AI. You know, it's a huge open new territory, new frontier. Everybody's jumping in, I think in the next few years, at least initially, we'll see accelerated development of AI technology to become more powerful. You know, we're touching on many other aspects of our lives, but this trends eventually will reach a saturation point because I also understand the current model, as powerful as they are, they're not, you know, you know, infinitely powerful. So at some point in time we'll see a saturation point. And also we, as we have better understanding of the AI technology, there will be more effective, I'll say guidelines and regulations. There are like, you know, the Wild West slowly becomes into a law-abiding community. I think that's going to happen. Now in terms of, you know, AI taking away jobs, I think that's for sure. Is this the same thing we have seen during the Industrial Revolution? Right, during the Information Revolution, we have... the typewriters, right? I mean, computers have replaced typists. I mean, we used to have people typing in all the documents by hand. And those jobs are completely eliminated by new technology, but it doesn't mean that the economy is going to collapse. We, we humans, just as I said, very versatile, will find new opportunities and we may... ride on the waves of new technology, creating new opportunities, new jobs, and cope with the situation. And that's where I see the most amazing part of human nature is being very creative, being original and finding ways. You know, you initially unthought of. So I think that's where we're going to see more of this sparking moments of human genius than, you know, in the past. So I'm, actually in that sense, I'm more optimistic about the future than pessimistic.

[Kevin]: Well, that's a great place to leave it, Siwei. So we will leave it there.

[Siwei]: Okay.

[Kevin]: I'd love you to come back some time and talk about some of the risks associated with AI, including deepfakes, which I know is one of your areas of specialty. Would you come back and talk to us about that sometime?

[Siwei]: Absolutely. Absolutely. Yes.

[Kevin]: Thank you so much, Professor Siwei Lyu of the State University of New York at Buffalo. Thank you so much for joining us.

[Siwei]: Thank you so much for having me. Yes, have a good day.

[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. Thanks for listening.

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.