Delivering Next-Generation AI Capabilities

Featuring Matt Gaston as Interviewed by Matt Butkovic

Welcome to the SEI Podcast Series, a production of the Carnegie Mellon University Software Engineering Institute. The SEI is a federally funded research and development center sponsored by the U.S. Department of Defense. A transcript of today's podcast is posted on the SEI website at sei.cmu.edu/podcasts.

Matt Butkovic: Hello and welcome to the SEI Podcast Series. My name is Matthew Butkovic. I am the technical director for Cyber Risk and Resilience in the CERT Division of the Software Engineering Institute. Today, I am joined by my colleague, Matt Gaston. Matt is the director of the SEI's Artificial Intelligence (AI) Division, where he leads a team of researchers, engineers, and innovators who assist the defense and national security communities in developing and adopting leap-ahead AI capabilities that are scalable, reliable, and assured. Matt has been working at the intersection of national security and technology for more than 25 years including serving as an officer in the United States Air Force and spending 10 years at the National Security Agency. In addition to his work at the SEI, Matt is an adjunct associate professor in the Software and Societal Systems Department in the School of Computer Science at Carnegie Mellon University. Welcome, Matt.

Matt Gaston: Thanks, Matt. I'm really excited for our conversation.
Matt Butkovic: I am as well. Matt, AI is undeniably this transformative force

in just about every facet of our lives to include national security. You have been someone that has witnessed the evolution of these technologies. Now you are leading efforts to harness these technologies for maximum benefit for our nation. I was wondering if you could just describe briefly what the division does and your position in relation to that mission.

Matt Gaston: Thanks again. I am really excited for this conversation. As you said, AI is transformational. It is changing the way we live, the way we work, the way we defend our country. It is important to look at the history of the Software Engineering Institute here at Carnegie Mellon University to understand what we do when it comes to AI. So, 40 years ago, the Department of Defense realized that there was this really important building material called software that was going to drive a lot of capability, a lot of innovation, a lot of technology into the future. They needed a place that could worry about how to do software as well as software can be done. For 40 years we have been doing that. Of course, if you are worried about how to do software right, you are worried about how to do cybersecurity right. So we have been doing that for a long time, longer than it has been fashionable. Now with the transformation that that AI is bringing, we are really focused on what we call AI engineering. It is not necessarily the latest and greatest, shiny object in AI, the shiny new technology, but it is how to use those technologies, how to build those technologies, how to adopt those technologies, how to know when those technologies will work, when they won't work, and what to do about it. We are really focused on what we call advancing and promoting the discipline of AI engineering. Of course, we can't do that alone. One of the huge benefits of being here at the Software Engineering Institute is that we are part of Carnegie Mellon University [CMU]. CMU is the birthplace of AI. On a recent call to faculty, I learned that over 300 faculty here at the university are working on AI—it is probably even more than that today—but also in partnership with lots of other organizations, other federally funded research and development centers (FFRDCs) in the DoD ecosystem. But even beyond that, to national labs, to industry. I would say that last part, the industry part is really increasing, because they are doing a lot of the innovation in AI and the Department of Defense, the national security community, can learn from those companies but also adopt those innovations for game-changing capabilities.

Matt Butkovic: Thanks, Matt. There is a lot to unpack there. Thank you for the description of what the division does. You described applying the

principles of engineering—engineering as a discipline to the topic of AI—and that seems at the center of the work that the division does. As you mentioned the *Software Engineering Institute* is in our name, right? We applied a similar approach to the foundational elements of software engineering. Can you compare and contrast the challenges you have in AI engineering versus traditional or software engineering?

Matt Gaston: There is a lot to unpack here. To start the conversation, I will point back to 2019. We have been worried about AI engineering since before the large language model and the now agentic AI craze that is happening for good reason. Back in 2019, we put out a document that was actually written by all parts of the SEI. We had the traditional software engineering perspective. We had the cybersecurity perspective, and we had the AI perspective, called 11 Foundational Principles. And principle zero—because we are computer scientists here, we start counting with zero—principle zero was, *Remember that these systems, these AI systems, are software systems*. So practice good software engineering, practice good safety engineering practice, practice good systems engineering, practice good cybersecurity engineering when you are doing AI. But then the rest of that document and the work that we have done since then is all about what are the special aspects of AI that really push or challenge these disciplines.

Remember, at least from my perspective, that engineering is really about understanding a theory of failure and understanding how to mitigate and avoid those failure modes in any kind of system. A lot of our work is really about doing that. It turns out that AI systems—and when I say AI systems, I am really talking about deep learning work from the late 2010s and now into foundation models, large language models and even into agents. We don't really have a theory of failure for these systems. In fact, as recently as last week, I was reading articles about how the companies that are building these amazing models that everyone is using still don't fully understand exactly how they work. They are investing to figure out what is going on inside of them. What are the patterns of quote unquote *reasoning* that these models have? And so ways that AI specifically challenges traditional engineering disciplines [are] really driven by data. Machine learning is the process of using a model, using an algorithm, to process data to find patterns in that data, to then build a model to make other decisions or generate new content. When it is data that is driving that functionality rather than a human writing down the rules, aka writing software. Rather than a human writing down those rules, the model or the algorithm is actually learning patterns that it is then to affect things in the world. That indirection makes them very hard to understand. That is one example. And you may have a question.

Matt Butkovic: Yes. I really like the description you offered. It seems what I am hearing is that the opacity or the inscrutable nature seemingly of some of these systems is leading us to need to find new techniques that weren't necessary in other traditional engineering disciplines. You describe fault modes or failure modes. It seems like there is an extra challenge of determining what those may be in something that has varied outcomes. And the data feeding the system rapidly resets the traditional engineering expectations you might have. Is that fair?

Matt Gaston: Yes, it is fair. I would say, if I were to boil it down, in traditional systems, a human or a team of humans, write down the rules for how they want the system to behave. It could be math for a traditional civil engineering project, or it could be code in a software engineering project. In AI engineering, you are letting these algorithms sort of figure out the functionality. When they do that, we don't have direct insight into what is going on. We don't have direct insight into all of the possible correlations. They are also in massively high-dimensional spaces, so it is hard to sort of introspect them. There are these well-known problems in AI called interpretability and explainability. We are still working on them, even in 2025. We are still working really hard to make progress on understanding how these systems are working from the inside.

Matt Butkovic: Matt, you mentioned, the element of this technology coming from industry or private industry into defense spaces and national security spaces, which is slightly different than maybe other technological innovations or evolutions that we have seen. There is a motto in some tech circles to go fast and break stuff and fail early. That is great in some contexts, but given the very specialized mission space that we support here at the SEI, thoughts about how that partnership with industry can lead to us having more trustworthy AI and perhaps avoiding negative outcomes in the deployment of these systems.

Matt Gaston: Sure. So again, lots to unpack there. Certainly, as part of the defense and national security ecosystem, I don't think go fast and break things always makes sense. I would say go fast and try things and understand what is working and what is not working. But I think you do point to a big thing that is happening in the economy in the United States, maybe around the world. And that is that traditionally several decades ago, DoD was really an inventor or a primary investor in new technology. DoD certainly still does that, but we are seeing massive amounts of private equity and venture capital and corporate dollars go into amazing innovations, certainly the

family of frontier models. I am not allowed to be biased and have a favorite, so I will just name a few: Claude from Anthropic, GPT from OpenAI, Gemini from Google. There is a list that continues. These models have been invented in the private sector. They have been invested in through private equity and venture capital, and the DoD is now sort of an importer of these things. I think that is a really important change, because the DoD, the national security community, has to be able to understand these technologies and bring them in and apply them in their domain in ways that they can count on them and trust them. There is a lot more we can discuss there, but I will let you lead the way.

Matt Butkovic: Thanks, Matt. Yes, to me, this is one of the most important takeaways, perhaps, of this discussion is that this is inherently different in the sense that the degree of acquisition-rather-than-construction of these systems that is going on is reshaping the way that the DoD and national security communities have to view this problem. That is why I think, what I know, is that the SEI is such an important partner in all this because we have occupied this space for many years successfully, where we guide the DoD and national security missions to the right answer or the most prudent answer when it comes to acquiring and operating these systems. I know that is one of the central thrust of the division is ensuring that we build that into the engineering disciplines. I think, broadly, there is a convergence. If the topic weren't AI, but it was storage or compute. If I look at the cloud, I think the DoD and private industry are more intertwined than ever before. And the defense ecosystem, the defense industrial base, as we traditionally describe it, is being reshaped again by AI. I don't think you can credibly say that AI companies aren't part of the DIB [defense industrial base] at this point. I know that was a bit of an opinion rather than a question, but how best do you think we can assist players of all sizes in the defense industrial base with the acquisition and deployment of reliable and robust AI?

Matt Gaston: Yes, you point out why the SEI is a good place for us to worry about AI engineering and helping the defense and national security community adopt these technologies. It is because we have had an acquisition focus, right? Software is largely written by companies, and the DoD acquires that software. You pointed to cloud and storage, those sorts of things. Yes. We have been helping the DoD do that for a long time. I think what is different about AI is the complexity of the capability of these systems that we don't yet even fully understand, but you can do a lot with them. You can mistakenly use them in ways that it is unintended. You can also work really hard to set it up so that you do it right, and you can rely on these systems, and that is really our niche. We do think it is critically important to

help accelerate the adoption of these technologies. One thing I like to focus on, our team is having our folks be what I call expert first users of these technologies for DoD. As an FFRDC, we sort of live between the commercial world and folks in the government, warfighters. We can show them what is possible with these technologies. We can show them the right ways to use these technologies. We can show them how to integrate these technologies into bigger systems with legacy capability or new capability that they are developing.

Matt Butkovic: Which is just imperative. There is no version of this where AI doesn't become part of the lifeblood of these missions, right? It is unavoidable. Now, at the national level, there has been focus applied to a few elements that we believe will best serve us to ensure that AI systems are trustworthy, reliable, if not completely explainable. One of those is constructing these ecosystems of evaluation. This is a space that the SEI has occupied for quite some time. In my work here, that includes testing the security and trustworthiness of networks. An extension of that concept is then doing red teaming for AI systems and other related AI elements. Thoughts, Matt, about how best we can construct these ecosystems of evaluation, not just simply for security but other attributes of system performance. how best we can construct these ecosystems of evaluation, not just simply for security but other attributes of system performance.

Matt Gaston: This is a great question and indeed a special area of focus for the SEI. Again, building out our legacy of testing software systems and cyber testing. Testing of AI systems is new and different and introduces new complexities. I would say one area of ours that we have developed our expertise in is test and evaluation. Just to give a glimpse into that, in the AI world, most people focus on performance. They focus on how well does an AI model perform at a given task. They don't worry about other quality attributes, to use a software engineering term, that that system might exhibit or have behavior that that they should focus on, like say, how confident is the model? If you are trying to detect a potential threat on say a network with an AI system, you would want to know exactly how confident the prediction is of a potential threat or potential normal activity in that network. What I mean by confidence is, *Does my prediction actually hold up*? If I am 50 percent sure, you wouldn't think I am too serious about it, but if I am 99 percent sure, it matters. And so, confidence, understanding confidence, calibrating confidence, quantifying uncertainty, these are sort of extensions to the typical evaluation around accuracy that we see. We specialize in that. We have helped big government programs, some of the leading AI programs in the DoD adopt this, this way of evaluating not just at development time, but

during, operations. Other parts of test and evaluation that I think are really important: AI does, in fact, introduce new security vulnerabilities to systems. There are special ways that AI models can be manipulated or deceived or evaded, or special ways they can give up information about what is behind them. And so we have to extend cybersecurity practices to account for these new types of vulnerabilities. And that is new classes of testing when we are getting ready to either use or deploy this type of technology.

Then the last one, and this is one I have been thinking a whole lot about recently, is benchmarks. There are hundreds of benchmarks out there that all of the frontier models regularly run against to evaluate their performance. It is pretty well known that those benchmarks are…They are useful, but they are often gamed by the system. And they are not sufficient for really, truly understanding the performance and capability beyond those benchmarks of those systems. I think one space that we must play in as a DoD FFRDC is helping the defense and national security community establish benchmarks that evaluate these tools for special defense and national security applications. There are special needs there: mission-critical, safety-critical, life-critical applications. How do we truly evaluate in those spaces, beyond just what is out in the marketplace for benchmarks?

Matt Butkovic: This is something that we have done traditionally at the SEI. As you pointed out, this is such an important intersection of academia and the Department of Defense. So I think we are well positioned to do that. One of the things we like to highlight, Matt, is transition. As an FFRDC, this is one of our primary goals. Is there a specific application of AI or a specific tool that you would like to highlight as a recent success story?

Matt Gaston: Yes. Transition is a complicated one. We often think of transition or technology transition as handing over a piece of software or a piece of software capability to a user to then use in their operations. But I think transition for us is very, very complex and has lots of different facets. My favorite story, I'll tell it. This is from several years ago. We partnered with the Defense Innovation Unit to run a challenge called Xview2. Xview2 is all about identifying damaged buildings after natural disasters from satellite imagery using machine learning models. We created a first-of-its-kind data set. We created a test-and-evaluation rubric for that problem domain, and then we opened the challenge to the world. This yielded all kinds of different implementations, I think over 1,500 submissions to that challenge. And the models that turned out to be the best, we have worked with the DIU over the last five-or-six years to integrate them in various application domains

including bushfires in Australia, wildfires in California, earthquakes in Turkey and Syria. From a transition perspective, not only did we help create these technologies, but we then worked very closely with our government partners to get these technologies into the hands of folks where they could really make a difference. If you think about that, understanding damaged buildings after a natural disaster, the typical way of doing that is probably send people in to do the evaluation, which is quite dangerous. Maybe send drones in, but you still got people within proximity. Being able to do this from, sort of a standoff capability with satellite imagery is a lifesaving capability. We are very proud of that work.

Now, I started by saying transitions are pretty complicated for us. Going back to the beginning, we think our mission is to advance and promote AI engineering. So how do you transition that? We can actually transition lessons learned from industry. We can go out and talk to the best folks out there in the world that are building AI systems and learn their patterns and practices, accumulate that up into the engineering discipline, and then share that discipline back with the defense innovation base, industry, academia, other places. A lot of our transition work is getting information, knowledge, best practices to flow across different aspects of the world that we play in to include transitioning ideas from industry to government, but also government requirements out to industry. Bill Scherlis, one of our senior advisers here, often talks about reality transfer. Actually, sharing the real problems that we care about with industry can spur innovation, can spur them to go off and make investments into solving those problems. We think of transition in this very complex way, which I think is an important part of it.

Matt Butkovic: Absolutely. I think that is the only way to look at it. These innovation cycles are getting shorter and shorter. Expectations are growing for return on investment in these technologies. As you pointed out, that is not simply financial return, it is also life safety and natural defense return. This is an extremely exciting time. You are here for a foundational change. I am as well watching from element of the organization. If folks are interested in things you described today, I know you have a catalog of resources on the SEI website or anything. Where would you suggest people start if they have a nascent interest in the topics we touched on today?

Matt: There are things that are out there. You can go to the SEI website and find all kinds of information about different aspects of AI and AI engineering and linkages to software engineering and cybersecurity. I really like our *11 Foundational Practices*. We recently did an analysis of those. Although we wrote them in 2019 and much has changed since 2019, we have revalidated

that they are still relevant and useful for people both getting started in thinking about how to bring AI into their environment or also seasoned developers of AI.

Another thing we are doing, and we are right in the middle of it right now, is a national study. I talked about transition and learning from industry. Our national study is really about the state of AI engineering and identifying important future directions for work on advancing AI engineering should focus. We are talking to lots of folks. We are talking to frontier AI labs. We are talking to other research labs. We are talking to big industry players in space, academics, lots of folks. That study should be published early 2026. We are in the midst of it right now. I am getting way ahead of myself. I'll share one finding that I think is pretty important. There is a discipline called safety engineering. That came out of MIT. The textbooks actually come out of MIT, and there is a, a movement in, in the AI community and sort of the practitioner AI community to start to think about how do we apply safety engineering in the process of designing AI systems? Not to be confused with sort of the big AI safety concerns, which we won't get into, but safety engineering in a very practical level. How do I understand edge cases? How do I evaluate in those edge cases? How do I bound the functionality of.

Matt Butkovic: Assurance cases, that sort of level of…

Matt Gaston: All of that sort of thing. The finding is that AI challenges traditional safety engineering, and AI needs better safety engineering. And so we are going to make some recommendations along how to make that happen.

Matt Butkovic: Excellent. For folks in the audience that are maybe only now contemplating how AI fits in their way of working in their organization, or maybe they have some apprehension about AI. Any thoughts about on how they can stop worrying and learn to love AI?

Matt Gaston: Well, I don't know that you should learn to love AI but maybe consider it a potentially very useful tool. I got in trouble at a conference recently. It was a DoD conference where I had a very similar question. *What do we do about all our worries about AI?* And I said*, Well, how many people have actually gone out and tried it*, *tried using it for some mundane activity like, meal planning or, or vacation planning or something like that?* At the time, this was maybe 18 months ago, there was only about 10 percent that raised their hand that they had tried these tools. I recommend highly that in safe environments for tasks that don't really matter, maybe un-work related, go

try these tools out. See what they can do. See how they can help you think through an idea. Notice I said, *Help you think through the idea*. They are really powerful. I use them in my daily workflow. My kids in college use them in their daily workflow, not for cheating, of course, but just to help them think through problems, understand different aspects, different perspectives. There are all sorts of interesting ways that these models, these large language models and chat bots that are out there now can help us.

Matt Butkovic: Excellent. One of the frequent items we see in the feedback is, *I am a student, and I want to learn more about the subject. I am interested in working in this field*. Any tips Matt? Let's say that you are an undergraduate, and you want to work in this very specialized niche of AI for national defense. Thoughts about things you should study or experiences you should obtain before working in this field.

Matt Gaston: This is a fascinating question because a lot of people think they need to go study computer science and get an advanced degree, maybe a master's degree or a PhD in machine learning, to be relevant in this community. I think that is actually changing. Those folks are the folks that are really going to invent new models, new algorithms, really push the state of the field, where I think there is a whole new set of jobs and opportunities on the use of AI. That can be incorporated through working in other disciplines. It can be incorporated through adopting these technologies in your field, in your industry. There are lots of different ways to do this. I suspect that in a lot of places, a lot of universities, a lot of maybe even trade schools, there are going to be learning opportunities on the implications of AI for whatever it is: for being an electrician, for being a physicist. That is already happening for sure. AI is being applied in science, but I think we are going to start to train this. One really interesting thing here at CMU. About three or four years ago, the College of Engineering decided they were going to be an AI-first engineering college. Bill Sanders, who was the dean at the time—he is now I think, chancellor at Rochester Institute of Technology or president at Rochester Institute of Technology. He actually created 5 or 7 new master's degrees in AI for, pick your engineering discipline, so AI applied in mechanical engineering. That is a really interesting way to think about this. You don't have to be the deep expert in the algorithms of AI, but you can be really an expert in how do I use AI technologies, not just generative AI, but also the good old-fashioned predictive AI from 2017. How do I use these technologies in my discipline?

Matt Butkovic: It is really important to understand, right? There are folks, like the folks in your division that are focused on AI engineering and then

creating the next generation of AI. But we shouldn't lose sight of the fact that for the majority of folks that touch AI, they will simply be users of AI. It will be augmenting what they do. You don't have to be an electrical engineer to use a power drill, right? I think it is sort of like that in my mind that is something sometimes lost, but I think that is natural when something is so new and so transformative. But, Matt, I really appreciated the discussion today. I was wondering if you had any questions for me.

Matt Gaston: Yes, I love to ask interviewers questions. Thanks for that opportunity. I think the most obvious question is you have been working in cyber risk and resilience for at least 15 years here at the SEI. Maybe before that in your career. With all of the changes happening in the last three to four years in AI, how has your outlook changed for cyber risk and resilience?

Matt Butkovic: It has changed in several important ways. So now AI is part of the things you have to contemplate when you think about things like attack surface. On one hand, we have a new set of security considerations that potentially make you more vulnerable. I would argue that extending what we know about how best to manage that in traditional computing is a leg up. On the other hand, as we have touched on a few times and did a good treatment of in this discussion, it is also an enabler. We can now do types of evaluation, types of analysis. Think about the risk calculations you can do that weren't possible without augmentation by AI. I think it is both increasing the attack surface, but all new technology does that. We shouldn't get too worked up about that. But it has tremendous potential. We are already seeing this to then make our, our efforts faster, better, cheaper and more reliable.

Matt Gaston: I think it is really interesting that, in the two most prominent applications of AI out there in the world right now, are in software engineering for code generation and in cyber defense. It is interesting that here at the SEI, we have those three areas all intertwined and working together.

Matt Butkovic: Yes, absolutely. That is a great point, which is one of the things that makes the SEI very special, is we have this long and well-established set of resources on campus to draw on, and also the ability to work across these silos and do something in with a unity of effort. Matt, I want to thank you for taking time today to talk about your work, your division, and where AI engineering is headed as a discipline.

To our listeners, thank you for joining us today. We will include links in our

transcripts to all the resources that were mentioned. The SEI podcast series is available in all the places you find podcasts, Apple Podcasts, SoundCloud, Spotify, and the SEI's own YouTube channel. As always, if you have questions, please don't hesitate to email us at info@sei.cmu.edu. Thanks again Matt.

Matt: Thank you, Matt.

Thanks for joining us. This episode is available where you download podcasts, including SoundCloud, TuneIn radio, and Apple podcasts. It is also available on the SEI website at sei.cmu.edu/podcasts and the SEI's YouTube channel. This copyrighted work is made available through the Software Engineering Institute, a federally funded research and development center sponsored by the U.S. Department of Defense. For more information about the SEI and this work, please visit www.sei.cmu.edu. As always, if you have any questions, please don't hesitate to e-mail us at info@sei.cmu.edu. Thank you.