Evaluating Large Language Models for Cybersecurity Tasks: Challenges and Best Practices

Featuring Jeff Gennari and Sam Perl as Interviewed by Suzanne Miller

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.

Suzanne Miller: Hi, and welcome to the SEI Podcast Series. My name is Suzanne Miller, and I am a principal researcher in the SEI Software Solutions Division. Today, I am joined by Jeff Gennari and Sam Perl, who are joining me today to discuss the recent collaboration with OpenAI that resulted in the development of 14 recommendations for evaluating large language models (LLMs) for cybersecurity tasks. Welcome, Jeff and Sam.

Jeff Gennari: Thank you.

Sam Perl: Thank you.

Suzanne: Good to have you here. You have both been guests on our podcast series before, but not everybody is going to know you. For those members of the audience who have not seen your previous podcasts—and we will link to

those in our transcript, by the way—let's start off by having each of you tell us a little bit about yourself, what brought you to the SEI, and tell us a little bit about the work you do, and what is the coolest part of your job? And why don't we go ahead and start with Sam.

Sam: Sure, thank you. For me, my journey to the SEI was a bit of a long one. I started out as an undergraduate and a graduate student at Carnegie Mellon University. And in my graduate student days, some of my professors were from the CERT department and other parts of the SEI. And I was very sort of engaged with cybersecurity and the courses that I was taking and really impressed with the faculty. I went off and did some consulting in the private sector for about nine years, not related to the SEI but still working in cybersecurity. Then somewhat randomly one day, got the opportunity to interview with the SEI and CERT. And so again, I had moved back to the Pittsburgh area. And I was just thrilled to be able to join that team. I was able to transition into a more research role, looking broadly at problems that would affect multiple industries and people and consumers and really excited and happy to be here and just have had a wonderful career here. I have been here close to 13 years at this point, lots of great research in the areas of incident response and cybersecurity in general as well.

Suzanne: Well, and we are happy to have you here, I have to say. Jeff?

Jeff: Thank you. Yes. My journey at the SEI began about 20 years ago actually. I was fortunate enough to get a job in an operational context at CERT doing computer security incident response, vulnerability remediation. As I have been here, we have evolved to become more research oriented, tackling different aspects of cybersecurity, doing a lot of engagement with Carnegie Mellon proper. I have been thrilled to see how the SEI has continued to evolve and adapt into different aspects of software engineering and in particular cybersecurity. I think the work we are going to talk about really touches on some of that or is really evidence of that evolution.

Suzanne: All right. Thank you, both. Today, we are here to talk about recent research efforts to evaluate the use of large language models for cybersecurity tasks. Can we start that by exploring the current landscape for AI [artificial intelligence] in cyber? Beyond cogeneration and evaluation, what applications are available for large language models in the cyber landscape?

Jeff: I think as you asked the question and I think about the answer, I think the possibilities are quite large. And I think that is part of the reason we

wanted to write the paper in the first place was every problem that we have in cybersecurity, vulnerability exploitation, vulnerability discovery, secure coding, risk management policy concerns can be impacted by large language models, AI chatbots in particular. I really think that just about anything has the opportunity to be impacted by this space. The space is also I think changing how we think about cybersecurity. It is opening up new attack vectors. The attack surface of our systems is changing when they have AI components. I think when we started this work, it was really about, *How do we use this technology in different parts of cybersecurity effectively*? You see evaluations for individual areas. How well does an LLM or an AI-powered assistant do a particular thing like diagnose a vulnerability or suggest a fix or something like that. But we really do not know how it will work in a general sense. What are the limits? What are the risks and tradeoffs? It is a long way to answer that I think that everything is in play.

Suzanne: Well, that is what I was actually going to say, is what you are really saying is there is a very large landscape here. It is not a small landscape in terms of how LLMs and AI in general either contribute to vulnerabilities or contribute to defense against those vulnerabilities. This is really important work. Sam, did you want to add anything to that?

Sam: I thought that was great. I fully agree. I am thinking, *Boy, we are going to need a whole new podcast just to talk about that one question*, right?

Suzanne: Well, we can do that.

Sam: Yes, just because there are so many, in my opinion, emerging uses of AI in cybersecurity. The other interesting thing which I would say is somewhat surprising to some people, although usually not the insiders, is AI has been, or sort of machine learning, has been used in cybersecurity quite extensively for a very long history. Spam filters have been using machine learning for decades. Intrusion detection systems were originated using statistical models of traffic and traffic analysis and classification. We are talking about much more advanced sort of techniques in those areas and what will come next and how will it be used. In my opinion, it is a very large field. And we are already seeing lots of extensions to existing platforms or existing tools that are now incorporating generative AI technology to allow them to be more easily used by their existing users or to attract new users or to, in some cases, try to replace users entirely. Getting fully autonomous. A lot of it breaks down into what kind of task are you trying to do and whether or not an LLM sort of powered model is the right one or do you want to choose a different kind of AI model, of which there are many. I think there is an

interesting research area, which is one of the fun things about being right at our university where we have both the academic departments working on building the new models and then our institute looking at *How do we use these and apply these correctly—appropriately and correctly—in security*? Essentially, we get the best of both worlds to play in that research sandbox. It is really an amazing place to be.

Suzanne: There is one aspect of LLMs that we have heard about in the press, and I am not sure everyone that is listening may know what they are and what their effect is. This is LLM hallucinations. And can you talk about sort of what are they, and what are the ramifications of those in particular for the cybersecurity domain?

Jeff: Sure. I think it is a simple concept where you ask a LLM a question and it provides an answer and some percentage of those answers are going to be just inaccurate. It is going to make things up, a lot of that has to do with the nature of the technology itself and what is actually happening behind the scenes. My personal experience working with LLMs in the security context is that the real challenge is not that the LLM is wrong. It is that it is very hard to detect the wrongness because it is one wrong thing in a mostly right answer. They are actually very good at giving you approximate answers. At least, this is anecdotal, I am sure there are numbers to back this up. Sam probably knows them. But the answers are generally, they sound right to an expert eye but a little bit of it is wrong. And it is that little bit and the difficulty in detecting the little bit that can be really insidious and really make it harder to build an overall trust that the system's doing what you expect it to be doing.

Suzanne: Sam, did you want to add anything?

Sam: No, I think that was a great answer. I think hallucinations tend to get a lot of attention from people but just like any other kind of tool, maybe you should not expect things to be entirely correct all of the time. Certainly, working with humans, we do not have that expectation that people always know everything and that they can also cite to the exact source of the original kind of thought that they have had. Yes, it is complicated though because we, in some contexts, have the expectation that if you are going to submit something or if you are going to provide a written answer—and I tell my students this all the time—if you are going to provide a written answer and the answer is lifted or partially informed by another source, you need to cite that source. And LLMs have a pretty hard time with that. And again, as Jeff said, the nature of the technology kind of in the back end and the way in which they were designed leads to sometimes they make things up. And it is

actually not that strange if you think about it, because if you are going to take a lot of the text on the Internet and use it to train something, it turns out that a lot of the text on the Internet is also made up.

Suzanne: Sure.

Sam: We write fictional stories all the time. We write thoughts which may or may not be true or entirely true. If we are going to basically just use most of that, it is very difficult to determine when this one is a fact and this one is not a fact in all cases. And in some cases, if you are a fiction writer and you are using an LLM, you do not want it to be giving you facts. You want it to be making things up in some cases. Yes, it is kind of like, *Well, wait a minute, wait a minute. What do we want to use this for? Which contexts do we want it to be factual and which ones do we not want it to be factual? And how do we enforce that and how do we verify it*? And I think that sort of leads us into, *How do we improve evaluations for different contexts, now that we have something which I think we have not had before at the scale in which it exists and an accessibility in which it exists?* Now we are trying to figure out what do we do about the evaluations for every different kind of context. And I think we were trying to make our small contribution for the context that we tend to work in, cybersecurity. What should we think about when we talk about how are these going to be used and how can we evaluate and verify and validate them?

Suzanne: That leads us to your blog post where you talked about four different categories and 14 different recommendations for evaluating LLMs specifically for cybersecurity tasks. And I am not sure how you want to do this, but can we give our readers an overview, at least an overview of the four sections and maybe some of the more interesting recommendations or more possibly controversial recommendations, if there are any, that came out of this research in terms of evaluating LLMs for cybersecurity? And just for our audience members who wanted to take a deeper dive, we will include links to both your blog post and the paper that it is based on in our transcripts. You do not have to take notes, this is not the only place you are going to hear about this stuff. Sam, do you want to go ahead and continue on and sort of get us started with this?

Sam: Sure, sure. I will just read all 14. That will not take long.

Suzanne: Yes. You know how we love that.

Sam: No. I think what our blog post was about and in fact, so we wrote a

paper which was a collaboration with the team at OpenAI. It was one of their research team members and their security team members. We did a lot of back and forth with them but ultimately came up with, I think trying to capture some of what we felt were existing gaps with what we had seen in how evaluations had been being performed and why we thought that maybe the cybersecurity domain was going to need a different kind of evaluation from what we had seen before. I think a lot of it was initially reactions to what we thought of as sort of an exam-based approach to evaluating what is this knowledge model, sort of language model, what knowledge the language model had picked up as it was being trained. That was what a lot of the evaluations were. It was like if we ask a series of questions and it can provide answers to those questions, then we have replicated the sort of capabilities of what a human would be able to do in that area at a low level. And I think in cybersecurity, even then we were trying to say, *Look, that is just not the case*. That is, number one, that is not the job. And number two, even if we use exams, because I give my students exams all the time, I know Jeff does too. If we use exams for humans, what is the actual implication of using that same type of exam for an LLM? It just does not fit. The same way, it does not give you the same kind of predictive power that an exam for a human gives you. What is the equivalent of trying to evaluate the knowledge that an LLM has in security, and for what kinds of security? That got us into sort of theories about security and the different knowledge domains. That is sort of just teeing it up. But a few of the things that we wanted to make sure we included were how do we get things away from the exam model and more towards like the real-world tasking or activities that practitioners would be performing. This ties into the problem. How big a problem are hallucinations? Well, we do not know, it depends on the task. If they are trying to come up with, the right command to input into, to discover if a vulnerability is present, there might need to be some kind of experimentation that can be done. There may not be an existing factual answer that has already been published, so we need them to try things that are somewhat new and novel. On the flip side, if there is a command that they should not run because of detection capabilities on the other side, then they are going to give the game away if they run that command. There are lots of nuances involved when it comes to these things. I started to get away from the answer a little bit at the end there, but I will probably turn it over to Jeff for articulating some of the practices that we talked about.

Jeff: Just to add to that, because I think the individual practices are in the blog post and I could read through them, but thinking back to the original idea of the collaboration with OpenAI was we have a lot of experience in cybersecurity, they are building ChatGPT and they are getting asked

questions about, *How do you evaluate this thing, this new creation, to make sure it is doing things in a secure way*. When we originally looked around to see how LLMs were being evaluated, it was the exam model. You would read articles, news stories about how ChatGPT can pass the SAT, for example. And that resonates with people because everyone has taken the SAT who went to college or at least are familiar with it. Saying that carries a lot of weight. But when we started to go down the path of what does that look like for security because we do have security examinations like CISSP, and those, they serve a purpose, but they do not make you an expert. They help serve as a checkpoint but really to be an expert or considered an expert, it is a bit more nuanced, there is an experiential component, you have to look at your accomplishments. The exam model just did not feel right because it was skewed towards the types of things that LLMs are good at, in particular factual recall. And we found that in security pretty much across the board, regardless of what specific area you are looking at, there are trade-offs that are made and risk assessments and nuanced reasoning that you are not going to get through a multiple-choice exam. A lot of the recommendations are from that frame. Designing the task so that it does not skew towards what the LLM is good at. Having tasks that are not just multiple-choice questions, true or false answers, they require a bit of nuance, there is some complexity. We are not saying that there is no role for that. You have to couch what you take away from the assessment appropriately, making evaluations robust. You want to avoid spurious results. You want to be careful in the data that you use to help set up the experiment. Really, you should think of this more of an experiment, less of an evaluation. You really want to design a task that you would maybe design for a human with some caveats so that you are prompting the LLM, no pun intended there, but you are trying to get a more reasonable reaction, not just a direct answer. The third area in the paper and the blog post was to properly frame your results. Do not over-extrapolate. Just because you do well on one evaluation does not mean you will do well on all of them. And to me when I look back on what we wrote and the journey we took in writing it, a lot of it just feels like common sense. I mean, it is a little eerie in how similar it is to teaching students at a university, but I think that is more in line with the nature of the technology.

Suzanne: I was going to say that metaphor I think actually works well because you can think of an LLM as being a student of whatever topic it is being asked to deal with. Just like a student, it is going out and researching what is available for me to use to answer the questions that I am being asked. Not all students perform equally, for one thing, and not all students access the right things, and not all students interpret things appropriately for

what they have been asked to do. But I particularly resonate with the idea you are talking about in terms of understanding what is the task in general that you are asking this LLM to do, because if that task is something that, if I do not know the general answer to the question that I am asking the LLM in cybersecurity or some other technical area, trying to say that they are going to give a better answer than I would, given I had access to the same resources, that is probably a fallacy. In my mind, the amount of trust that we put into these LLMs has to be balanced by understanding what their limitations are. And that I think the evaluations you are talking about address that. Am I on the right track with that?

Jeff: Yes, yes. Ultimately, we want to know how to best use this technology in cybersecurity for a variety of specific tasks. I think part of building trust in the technology can be helpful is evaluating it in a realistic way. Just having it pass a multiple-choice exam looks good on paper, but now you have to use it on the job in a possibly sensitive operation or task. And I would need a little bit more. It would not just be . . .

Suzanne: Oh, sorry.

Jeff: Oh, go on.

Suzanne: One of the things that Sam mentioned was citations. The ability of LLMs to provide the logic behind what it is that they have provided to you. Is that an area that LLMs are working on building up? Because I know in some of my own work with LLMs, that has been a frustration for me is, *Where did you get that from*? And when I ask for, *And what were your citations for that*? It is not always very satisfactory. Is that a particular area that for cybersecurity, we need to be careful about because if we do not know where they got the information, it could be, as a matter of fact, malware or some other kind of attack? Is that something we should be concerned about?

Sam: Yes, I think it is a complicated question with a complicated answer. I will . . .

Suzanne: I am good with that.

Sam: Yes. I will caveat it by saying, I think that the companies and developers who are working on generative AI models, particularly LLMs specifically, are trying to improve their ability to provide citations. With every version or new model that comes out, there is an attempt at improving that capability. I call it ongoing kind of research. And that is in general, that is across all different

kinds of contexts that they are being used in. And cybersecurity is not the only area that is dealing with that problem. There are other areas that are basically waiting to figure out how that problem is going to get solved before they really heavily deploy the use of LLMs for more serious forms of work. But that said, I mean, a lot can be done even today with the sort of limited ability to provide citations. Yes, it is an ongoing challenge. The other thing, I mean, security, again, one of the things we talked about a lot is even within security, there are a lot of different domains, vulnerability management, reverse engineering, penetration testing, policy and risk management, the usable security for human and computer interaction, all of these are really different from each other in many ways. And in some cases, you need citations, and in other cases, you may not. You may just need to get something that works. Yes, it is complicated even within cybersecurity. It is not cut and dry from one task to another. We talked a lot and including in the paper about how these different domains will also potentially need different evaluations as well. You cannot take a secure coding evaluation and use it for pen testing. It is just not the same task. That is part of my answer, too, is that I think even within these fields, using cybersecurity in particular is just the one we know best, it is going to be complicated to evaluate the performance of these tools even from one task to another. That is why we really focus on, *What are you trying to use this to actually do*? *And how are you evaluating its capabilities in the context of somebody who is very good at let's say reverse engineering, for example?*

Suzanne: Jeff, did you want to add anything into that?

Jeff: No, I think Sam captured everything quite well. I guess perhaps one thing that I find particularly challenging is that the pace of advancement, and this might be slowing perhaps a little bit, but the pace of advancement over the last two years has really been disorienting in trying to figure a lot of this out. I actually do commend a lot of the companies and researchers out there for at least recognizing these types of problems with their technology. I could think of analogs and other areas that took much longer to recognize security as something that should be taken seriously. Even just looking back at the paper which we published in the end of February [2024], middle of February, things are different now. I think that that pace is really hard to keep up with. And I am heartened by the fact that a lot of these organizations are recognizing that there is an opportunity and also some risk that they have to mitigate.

Suzanne: These are a couple of challenges. Are there any other particular challenges using LLMs and cybersecurity that you want our audience to be

aware of? We have talked about the pace. We have talked about the, I am going to say the breadth and diversity of the domain. And we have talked about sort of areas where there are particular known gaps in terms of what the LLMs in particular can do. But are there any other areas that you say, *Hey, if you are dealing with LLMs and you are concerned about cybersecurity, here are some things that you are going to have be up against*?

Jeff: There is one thing that is not really in scope for our paper but in AI security I think in general, and it is that it is just software. It is a remarkable technology, but it is still just software. And it has weaknesses and vulnerabilities itself. I think, evaluating this technology in my business, I would keep in mind that it is not magic and that there are some tradeoffs you have to make from your own cybersecurity perspective, not what the LLM will do for you. I am starting to hear that conversation a little bit. For example, when you put information into an LLM, the LLM generally has that information. There could be a privacy concern there. But other things like genuine run-of-the-mill software vulnerabilities. I am starting to hear a little bit more of that and I think that folks using this technology should keep that in mind that it is software. It has bugs and some of those bugs are security-relevant and can have bad consequences.

Suzanne: Right. Even the most wonderful magic has its problems. And you say it is not magic, it is just software. But I can tell you from at least my first encounters with ChatGPT, when I first heard about it, it felt like magic. I think that is part of the beauty but also part of the challenge, is it does feel like it is not just an evolutionary step, it is a revolutionary step in our ability to access information and get much more complex answers to complex questions. I think that aspect of it is back to your evaluation, understanding what it is you are evaluating and not just giving the exam version of the test, understanding the complexity of what you need to evaluate I think is one of the things that on the face of the LLMs, it looks like magic. You have got to get past that view and think about it, *This is just software, and how am I going to evaluate software and for cybersecurity and other things,* is going to be important to the evolution of this whole area as people are becoming more enamored of it, and to a certain extent more dependent on it. You said at the beginning, this is a very rich opportunity for research. But in terms of what we are talking about here, the cybersecurity aspects of LLMs, what are some of the transition things that we have available, the transition resources that we have available to people to help them to get more knowledgeable about this area and to provide them with some practical ideas about how to apply LLMs to cybersecurity tasks, and conversely, how to evaluate the cybersecurity of their LLMs? What have you got to offer for them?

Jeff: Are you talking about SEI specifically?

Suzanne: SEI and the general resources, not just SEI.

Jeff: Right. I think there is no lack of research in this area. Our own blog is full of specific posts applying LLM to different problem sets and domains, using it to help improve software engineering. There is lots of information that we put out specifically to support engineers and decision-makers. I think more broadly, it is the usual things. All these organizations, all these companies that are producing LLMs I think are trying their best to put out guidelines and guidance for proper use. Sam, I do not know if you have anything else specific you have encountered.

Sam: I would just say, because I have been trying to do my best to read as much as possible, and there is a lot coming out in terms of both direct results of attempting to apply LLMs to cybersecurity operations-related activities, as well as many of the other domains that we talked about that are in cybersecurity and that we mentioned in our blog post. I would definitely agree with Jeff. The SEI Blog has other groups at CERT and at the SEI and in our AI division making postings on using LLMs in different contexts, and then also in some cases evaluating the performance. I have definitely seen a lot of research of evaluating the performance against existing datasets or sometimes often new datasets that have to be collected just to perform the evaluation. And I think that is probably a direction that we are going to see a lot more of, is to put together some kind of dataset that typically is going to need to be very large and, in many cases, heavily curated and then used in some kind of evaluation. If you are looking at, let's say vulnerabilities or secure coding, we have a team that just focuses on secure coding and has been doing that for decades. They have been doing their own attempts to evaluations, and they say, *Listen, here is some code to tell me what the vulnerabilities are in it*. And then attempting to use LLMs to do that classification type of problem. But also, if you can find one, why not fix it? That is a new, somewhat new capability because now you can generate code. Can you generate code that is secure? That has been a big problem in the space of just general software engineering. And initial results from LLMs, at least in the first iterations of these models, has been that the code that they generated was not entirely secure, although I expect that to get better. As Jeff said, those companies are taking that pretty seriously. If you are trying to generate code and it is being generated insecurely and people are telling you that, you are listening. And you are doing what you can to fix it on the back end to say, *How do we improve these models so that we do generate something*,

what they do generate has more security? I think it is good business practice, too, to do that because your customers, that is what they want. Yes, I think it is really interesting wide-open space. I will go back to the blog post and say one of the most important kind of things that we wrote in there is to really be careful about overgeneralizing your claims. When you make a claim, because this technology as you said, it could feel like, *Wow, this is something we have not had before.* Turns out we ask a lot of questions, and we want complex answers. And this is getting a lot better at doing that. But once you start getting into those more niche areas, maybe the answers are not quite as accurate as the more generalized knowledge that you just find a lot more information about on the Internet. You get into other issues when you start getting narrower and narrower and narrower, which is what sort of some of the professional activities like in cybersecurity tend to get pretty narrow pretty quickly. I think I am excited to continue being able to do the research in this area. It is an exciting area to be in.

Suzanne: Going to that, what is next for you both on the research front? What are you going to be working on next and what can we talk to you about in a few months?

Jeff: You can go first, Sam.

Sam: I would say I have been doing some interesting kind of collaborations. I have actually been working with a team of lawyers over the past few years. One of the things we did was look at when the GPT-3 model came out, we had this collaboration where we said, *What would be the impact of something like this*? This was before ChatGPT. We wrote how that might have an impact on patent law. That was an interesting paper. And we have continued to collaborate on papers of that nature. What is the impact on law which uses a lot of language and often requires a lot of citations as well? And in my opinion, I think there are a lot of synergies between legal aspects and cybersecurity aspects of tasks. Stuff is within rules and you use a lot of precedent and experience and history. Yes, I am pretty excited about continuing to collaborate on what impact these technologies have on law along with the cybersecurity implications. But I have also been starting to look at some of the aspects of what does it mean to define things like fairness, what does it mean to define things like responsible? And then, my general area is often on the data of cybersecurity. In my case, it is more incident response. What do we do when we have an incident? How do we track it? How do we coordinate it? How do we reduce the amount of time that it takes our teams to recover and respond? Yes, I expect to be doing research in all of those areas, yes.

Suzanne: And for you, Jeff?

Jeff: My group is code analysis and reverse engineering. We have been building tools to assist reverse engineers, in particular malware and vulnerability analysts, to do their jobs better. And one area that I am excited to get into is reverse-engineering AI things, AI components in a variety of different contexts. When I say that it is all just software, I say that while spending my time trying to figure out how to take that software apart and find the good parts, so to speak. I mean, that is an emerging field I think in general. I am looking forward to seeing how that develops, seeing what we can learn from the artifacts and how we can use it to assist cybersecurity folks.

Suzanne: Excellent. Excellent, excellent. I really enjoy talking to you both, and I want to thank you for taking the time to talk with us about your work. I know for myself every time I engage with a large language model, one of my thoughts is, *And I wonder if there are any security issues related to what I am asking or related to what I am getting*. I am glad that somebody is paying close attention to this. And I did find your categories and your recommendations, I do agree with you, they are mostly common sense, but as we know in the world, common sense is not that common. I hope our readers will take to heart the recommendations that you have made in terms of understanding how we can evaluate these models for their use in cybersecurity. To our listeners, I want to thank you for joining us today. As always, we will include links in our transcripts to the resources we have mentioned in this podcast, so you will have lots of things to look at. If I were to be a little bit flip, I could say you could probably ChatGPT a question about, *What did we talk about today*? And that would be interesting to see once the podcast is published if we actually got real answers to that. Anyway, I want to thank you both for taking time to talk with us. And I want to thank our listeners for joining us, as well. We will include links in our transcript to all the resources mentioned in this podcast, including the blog post and the original paper. I also want to mention that Shing-hon Lau, another SEI colleague, is also another author of that blog post and the paper, who unfortunately could not join us today. The SEI podcast series is available in all the places that you can find podcasts, including Apple, SoundCloud, Spotify. And my favorite, the SEI's YouTube channel. As always, if you have any questions, please do not hesitate to email us at info@sei.cmu.edu. Thank you.

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 do not hesitate to email us at info@sei.cmu.edu. Thank you.