AI LEADERS PODCAST
EP 66 GEN AI AND DATA PRIVACY: A DIVE INTO CONFIDENTIAL COMPUTING
TRANSCRIPT

[00:00:00] **Rishabh Poddar** As an industry, we need to be thinking more holistically around adopting strategies so that we can adopt generative more responsibly.

[00:00:18] **Teresa Tung** Hi, welcome to another episode of our A.I. Leaders Podcast. My name is Teresa Tung. I'm Accenture's Global Data Capability Lead. I'm so pleased to introduce this podcast on data security in the Age of AI. So, we know data privacy and security have long been a concern for companies, but Generative A.I. takes it to a new level. Now every company feels a sense of urgency because to use generative AI one, they need to rethink what data security means. And they're recognizing there's new opportunities because of generative A.I., but there's also new emerging risks. So, I'm so pleased to be joined by two experts in the field. So maybe we'll start with Raluca at. Could you introduce yourself?

[00:01:04] **Raluca Ada Popa** Sure. I am happy to introduce myself. I'm Raluca Ada Popa and I'm a professor at UC Berkeley. Work in computer security, especially focused on the security and safety aspects of generative AI lately. I'm also a co-founder of Opaque Systems and president of Opaque Systems, where we try to enable Gen AI to run with confidential data.

[00:01:25] **Teresa Tung** Thanks Raluca. Rishabh?

[00:01:28] **Rishabh Poddar** Yes, thank you so much for having me, Teresa. I'm Rishabh Poddar. I'm the co-founder and CTO of Opaque Systems. Quick background. I did my Ph.D. at UC Berkeley. Raluca was my best advisor and together we did the research and built the technology that ultimately led to Opaque today. So, I'm super excited to be here.

[00:01:49] **Teresa Tung** Thank you both for joining. I'm also proud Berkeley alum. So go Bears!

[00:01:55] **Rishabh Poddar** Go Bears!

[00:01:56] **Raluca Ada Popa** Go bears!

[00:01:58] **Teresa Tung** To kick us off, right? So again, for organizations that are looking to use generative AI, they have a lot of data right across organizations, across boundaries, maybe even with third parties. So, this in turn creates new challenges and risks compared to even a year or two ago. So, the first question is for Raluca. What are some of the new risks that companies need to be aware of when using generative AI?

[00:02:22] **Raluca Ada Popa** Absolutely. There's a lot of new risks and some are exaggerated. I would say that the major risk is within the reality of proprietary and private data. If we send confidential prompts to an alarm provider, then the scrums could contain code and no IP data. And there have been a bunch of news that poll people have read on how a few companies are concerned about that and turned off the use of target internally, for example. There's also concern with confidential data when you fine tune those models on sensitive data or even when you train. And then the model remembers too much in many cases. Besides, there's concerns about integrity. What if the data was slightly poisoned and as a result of that can lead to predictions that are dangerous in some specific cases? Or how about prompt injection attacks where part of the prompt is generated adversarial and can make the model do something completely different and again, something dangerous or jailbreaking and the list goes on and on. It's a whole new space. So, safety and security are to a large extent, you know, open problems and in their infancy.

[00:03:36] **Teresa Tung** Yeah, I think it's certainly a new world out there. But I think some of the common patterns that you mentioned. Right. So, most people are prompting. Right. Very few people have started to build their own alarms or fine tune alarms that that'll come later. But even if you're using, as you mentioned, a third-party alarm, there have been some stories either during the prompting or when you're passing it to a third party, there's always a risk that people are now more sensitive to.

[00:04:03] **Raluca Ada Popa** Absolutely. And especially if you really want to use these alarms for increased productivity, then you actually do want to include in your prompt, you know, some information that could be sensitive because you do want to alert them to help you with respect to, you know, work on IP data and or confidential data or suggest things about user personal data to improve their lives. So, it's that tradeoff. On the one hand, you don't want to include confidential data. On the other hand, you are revealing it to the alarm provider.

[00:04:34] Teresa Tung Okay. So, stay tuned. The good news is we're going to come up with a solution for just that very common problem. But I wanted to also continue painting the landscape with Rishabh. So, in addition to generative AI, introducing new risks on using generative AI, but there's also some new risks that are made possible by scammers because they're using generative AI as a tool.

[00:04:58] Rishabh Poddar I couldn't agree more. Teresa So absolutely. I agree with the fact that companies look to adopt generative AI. They are dealing with a lot of anxiety because it presents new privacy risks, and they don't have a holistic framework to be thinking about that and they want to adopt generative AI. But what does that mean for data? What does that mean? Look, I mentioned when they're trading models, they need to make sure that the data remains because the confidential data does not leak. Should they instead be deploying models as part of the right pipeline? Are they allowed to reveal that data to the model provider? What happens when results are produced? Is a specific user allowed to see the results that they're getting? So as an industry, we need to be thinking more holistically around adopting strategies so that we can adopt generative more responsibly. At the same time, AI is getting weaponized, it is introducing new risks. If we take a step back and look at the state of data protection in the world today, right, most organizations rely on traditional data obfuscation methods for protecting their data.

They're anonymous in the data. They're hashing it, tokenizing it, redacting sensitive information before operationalizing it. Now, security researchers know that these techniques are not truly bulletproof. You can reverse anonymous data fairly well. You can reverse anonymous data somewhat easily if you have access to certain auxiliary data sets and sophisticated tools. But this problem has been further exacerbated as a result of generative AI, which has decreased significantly the barriers from mounting attacks that can reverse traditionally deployed obfuscation methods. I'll give you an example. There's something that we are researching, and we have been looking at Opaque as well, and we recently published a blog post about it. Let's say you have some anonymous data, right? And you think it's anonymous, maybe that some tokenized API feels no one can figure out whether this is Teresa has information on this, etc. You can take that anonymous data, throw it inside a ChatGPT prompt and ask: Tell me who these people are in the underlying data set. By providing the model with some auxiliary information in the form of publicly available datasets. It's as simple as here are some auxiliary information. Here is the anonymized data and tell me who the anonymous people are, and it will spit the response out. So, any Tom, Dick and Harry who has access to a chat bot can now take anonymous data and anonymous it very easily. So traditional techniques do not work anymore. They have been broken, but there weren't any solutions for it. That's the best that you can do. But this problem has become much, much worse as a result of general delay because it has decreased the barrier for attackers and anyone armed, the chat bot can now exploit it. Yes, this is just the tip of the iceberg.

[00:08:07] **Teresa Tung** In fact, Warren Buffett, he says that air scanning is the next big growth industry because it's so easy to do. But as I mentioned, there's good news, right? There's good news about both these problems. So, whether we're going to keep sensitive workloads confidential when using generative AI. I like what we're talking about. Or we're going to protect against malicious software and rogue actors. We got this new hope that you've been working on around gentle computing. So Raluca, can you explain this concept of confidential computing and its significance for data privacy?

[00:08:40] Raluca Ada Popa Sure, absolutely. These are clear. Let's take as an example the prompting scenario. Imagine the user sending an encrypted prompt to the provider. By encrypted, I mean, it's basically obfuscated, but in a way that's, you know, mathematically sound. We don't know how to break encryption. And then imagine and the provider has a way to perform the inference on the encrypted prompt without ever decrypting it. And in the process of doing so, all the intermediate values generated are also encrypted, and the final result is also encrypted and then sent back to the user who can decrypt. Now this is essentially the instruction that confidential computing provides, that on the provider side, the computation happens on encrypted data. The reason why this is nice is that the provider wants your prompt to order results. Also, hackers breaking in the system won't see the prompt or the results, but the user was still able to use the system like normal. Now confidential computing basically implements this abstraction with the help of specialized hardware.

This hardware deep down into the CPU use and the GPUs encrypts the data at all times before it leaves to memory on the bus. Now deep down in the CPU and GPU. It does get decrypted, but only the macro. Consider the decrypted data. For example, an administrative root access can only see encrypted data because they're looking at memory or the hypervisor can only see encrypted data. So basically, from the perspective of software running on that server, or you all employees or administrators of that service or hackers to that service, they can only see encrypted data. So, it's really powerful in that it provides a large degree of security against all kinds of actors that may look into what's going on the server side but also promises the service to run normally. And the good news is that this is actually very efficient. The hardware-based mechanisms are very efficient, there are cryptography alternatives, but the risk-based approaches are actually very, very efficient. There's minimal degradation to performance in some cases, almost not at all.

[00:10:47] **Teresa Tung** And where do you get this hardware from?

[00:10:50] **Raluca Ada Popa** Absolutely. So, the nice thing is that this hardware exists on all the major CPU vendors like AMD and Intel. And also, right now it exists on the latest iterations of and redistribute like the H 100 and is going to be on the block with architecture as well.

[00:11:09] **Teresa Tung** Yeah, and I think so. What I really like about this solution is its turnkey. You don't need to know the algorithm. You're running ahead of time. You don't even need to give too much thought about the use of data a priori. Right? You encrypt all the data and then you pick the environment. If I'm going to cloud provider, I pick just like I'd pick a VM that has a number of cores or amount of memory. I pick the one that has the special hardware, right, the secure enclave. And I think, is it as simple as that?

[00:11:39] **Raluca Ada Popa** Well, it's actually not as simple as that, but it's as simple as that in terms of not needing specialized hardware, in terms of not needing as a user of this technology to acquire special hardware or special servers, you can just run on the public clouds. And the public clouds already have this hardware, and you can use special VMs called confidential VMs. Now of course, if you want to have if you to deploy your ML pipeline in this or you want to do all this private prompting, if you just use the cloud directly, there's a lot of complications you have to think of. You need expertise, really. You need to think of how you handle keys. There are many cases involved here. There's a process called remote station in which the user can And is running the code that they're expecting, and that's also quite complex to set up. Then you might want to do a cluster computation. And I don't just want to use one machine. And then how do you expand the uncleared boundary to more machines and scale out? So, this is, I suppose, required expertise. And our work in our research at UC Berkeley aimed to tackle them, to make them very easy to use and all the way into the opaque product, which basically abstracts out all these difficulties. So again, the hardware is available on the cloud, but to use a platform like Opaque so you don't have to worry about, you know, how you scale out in key management and whatnot. And then it really is very easy to use because then you don't need any expertise. You can just run your workloads in or pick VMs and that's it.

[00:13:06] **Rishabh Poddar** I agree. I like to think of this in terms of two layers, right? You have the infrastructure, which is confidential computing hardware, and the infrastructure that the cloud providers provide that allows you to deploy workloads on that hardware. But you still need the ecosystem, the software ecosystem that can use that confidential computing capability in a secure and an end to end, secure and scalable way. And that is hard. It requires expertise and skills, like Raluca mentioned, how do you get data in properly? How do you verify that you're actually deploying this on secure machines? How do you do the scalability? How do you ensure that the workloads that are running on the machine are not actually leaking your data? It's simply not enough to keep the data encrypted in memory. If you deploy a query that says Select Start, I've suddenly got access to all of Theresa's data. So, you need to be able to enforce policies. You need to be able to verify that everything happened properly. You need verifiable audit trails. You need mechanisms and abstractions to deploy your workloads. Analytics, machine learning, whatever you want to do in a secure and scalable way. So, you still need that ecosystem of software components that makes confidential computing turnkey.

[00:14:20] **Teresa Tung** But from the perspective of an enterprise user, right? So, we need to make the decision as to which data to encrypt. We need to make decisions on the use cases. We need to make sure we get the hardware, possibly even using a cloud provider to make it easy. And we need to make sure all the hard stuff that you just mentioned is built into the software framework. So that is an investment. And what you've done research on is part of opaque, but that that software layer and you're almost like a very advanced model management capability but has the additional level of protection is pretty much in the software itself. So again, from the end user perspective, I don't have to handle some of that hard stuff. Right, right. That that's what you figured out as part of your research and your work. So, let's bring this to life Reshape. Can you talk about using this capability for generative and for things like ROMs and even fine tuning. But the same technology doesn't have to be so advanced. Right. I think a lot of the questions that enterprises have are just about being able to see more. Right. Having more data to be able to see a bigger picture. And I think sometimes it's not something that you have as a single party but being able to have multiple parties work together to see how you can bring to light how confidential computing enables also multiparty computing.

[00:15:49] **Rishabh Poddar** That's an excellent question, Teresa, and I'll maybe give you an example of a use case that we're working on with Accenture regarding telcos. What we are doing here, Accenture in a big are helping telcos to work together to identify IPS that are suspect or malicious. Now, a single telco working alone, they don't have the data that they need to see the complete picture. To see the complete picture, they need to be able to collaborate with other entities in that business ecosystem, other telcos and so forth, or other organizations. But they can't share their data with each other because the data is super confidential. There are laws and regulations in place that prevent this from being shared. They're often in competition with each other. So how can they work together to share intelligence, identify common threats, and take action against bad actors?.

A platform powered by confidential computing provides a way because now each telecom can keep its individual data encrypted using keys that they manage, but they can combine the encrypted data together and then run, analyzes, deploy machine learning or AI workloads on the combined data to get the insights that they want that benefits everyone. We are working on more complex use cases as well because that allows them to check more data points for malware detection, For example. By integrating more attributes, you can now build resources to build patterns that enhance threat detection for the entire industry. So confidential computing is a hammer essentially gives you the ability to maintain your data, privacy and sovereignty in a verifiable way. And this ability unlocks many use cases. You can now share data, combine data across different silos that can be both within an organization or across organizations, while ensuring that all data privacy requirements, governance regulations, they're all verifiably followed.

[00:17:56] **Raluca Ada Popa** It's a very powerful obstruction, a very powerful tool that really enables you to do a lot of things that you can do with confidential data. You know, we can protect the promise. You can clean up and sanitize the promise you can do, like Urrutia says. Collaborate across silos. Put together the insights from the data in training or analytics. It's horizontally very powerful.

[00:18:18] **Teresa Tung** I love the word verifiable way because one of the reasons why companies may not want to share data is it's hard to do. But oftentimes you're relying on a legal agreement and a promise which is important. But now you have this programmatically verifiable way that, yes, depending on each use case, I can see exactly which party contributed the data. I can even see which party might have contributed to the model. Right.

The algorithm that we work computing, which is a lot of IP as well. Right. So, I think that there by the way, when we see the results now, we can actually verify these are the contributors of the data, the contributors in the model. So, if we're thinking about the future, where data and models are really our competitive advantage as companies, this is really paving the way to enable data sharing with yes, even direct and indirect monetization.

[00:19:13] **Rishabh Poddar** I think this is also therefore, Teresa, in my view, a very effective tool to enable responsibly when we look at fairness and transparency and accountability, the ability to be able to verify and trace all data related actions and events that led to certain decisions is very powerful because now you don't have to trust anyone. You don't have to trust the people deploying the models or the people providing the software. The platform here is a tamper proof trail and a trample proof audit trail that anyone can dig, beat an auditor or a regulator to verify that yes, the data was only used in permissible and compliant ways and was kept secure at all times. So, I see it as a very powerful tool for responsibly as well as for compliance.

[00:20:08] **Teresa Tung** Still powerful even without Gen AI. Right. With data sharing across the board, across partners working with different models. Raluca, we want to go back to the use cases of compute, computational compute or generative AI implementations. Can you give us an example of something that we're that brings that example to life?

[00:20:28] **Raluca Ada Popa** Absolutely. Yeah. So, for example, again, in collaboration with Accenture, we're working with a major automotive manufacturer where we're basically using the confidential AI platform, mobile cake to take prompts and sanitize those prompts to remove confidential data. For example, remove prices, remove locations where the vehicle or customer identity was interested in what, but still replace this with meaningful symbols such as, you know, vehicle I one, you know, customer three things like this. But a limited understanding of the limb can still produce very useful answers, but the limb doesn't get to see the limb provider. It doesn't get to see all this PII data. So basically, inside the opaque, confidential A.I. platform, we are sanitizing and cleaning these prompts and replacing the PII with placeholders, sending data to the limb. And then on the way back we can be repopulated with the sensitive data so the user like before doesn't notice any difference. They don't need to be bothered by the fact that, you know, confidential care is protected at all. They have the same usual workflow was just starting with an alarm. So, they get all the benefits but at the same time protect the PII. And the nice thing is that, you know, once we have visibility inside the platform, once the customer has visibility into all the different user problems, they can also ensure compliance, like we said, can ensure, you know, accountability. Who asked what, who uses it for what and enforce policies across the board?

[00:22:02] **Teresa Tung** Yeah. So, this is 100% needed any time you're going to be using any sensitive data. So that's PII, that's IP. And especially if you're going to scale generative across the boundaries of an organization.

[00:22:17] **Raluca Ada Popa** And to add that importantly, here to FAQ and that all these sanitization and compliance tools and accountability, one inside a big platform, it means everyone, the Iranian citizen colleagues in LP doesn't get to see anything of what's going on. Again, because of the encryption, the cloud provider doesn't get to see employees, you know, internal employees to the cloud provider and provider or hackers like it really protects against such a large class of attackers here.

[00:22:48] **Teresa Tung** Everybody in the value chain is protected. So, I guess we talked about a lot of hardware, we talked about the software layer and all the things that it needs to do. So are somewhat needed so that we can have an end-to-end solution in place for confidential computing.

[00:23:04] **Rishabh Poddar** To confidential computing is a hammer, right? It ensures that your data is encrypted at all times. Just lifecycle to have an end-to-end solution, a more holistic ecosystem of components that utilize this capability in a secure and scalable way. I think Raluca touched upon this earlier as well, but how do you make sure that you are first deploying a confidential computing environment? You need to be able to test it and you need the requisite ecosystem infrastructure around it. How do you make sure that data is encrypted securely and is ingested into the platform of your choice in a scalable way? How do you deploy machine learning and AI workloads that take advantage of confidential computing that harnessed this capability in a secure and scalable way? How do you apply policies that enforce that? The data is only used in ways that you wanted to that you permitted to. How do you export audit trails that give you a log of all activities that took place on the platform? So, for confidential computing to be truly useful, you need a business application layer of sorts that harnesses this capability and makes sure that it is doing so in a secure and scalable way.

An analogy that I like to sometimes think of is you have hardware accelerators, right? GPUs just using a machine that has a GPU connected to it is not enough. You still need the right software abstractions and the right software tools that can use GPUs in a scalable way so that the user doesn't have to worry about it. You still need the GPU drivers; you still need the machine learning frameworks that are programed against GPUs to allow the developer that decided to take advantage of them. Similarly, in the confidential computing world at Opaque, our focus is on harnessing confidential computing for data and the AI workloads so that it's as simple as you want to run a machine learning workload or an analytics multiple choice connect your data sources, apply policies, develop your jobs using notebook or no interfaces, and then get the results. You don't have to worry about any of the complexities, the underlying comprehensive computing infrastructure, all of that is abstracted away from the user.

[00:25:27] **Teresa Tung** And so instead you're going to focus on the permissions that you want to take. You're going to focus on the use cases you want to create. You're going to focus on possibly monetizing data or using this for programmatic compliance and risk. And so, you can start more with the workload, not necessarily with the tech.

[00:25:46] **Rishabh Poddar** Exactly.

[00:25:48] **Teresa Tung** I do want to bring that back. So, we've been talking about encryption, and this is where a lot of this tech is based off what we're saying. The secret is because we're using encrypted data, and we can compute on the encrypted data without decrypting the data. So, one of the questions that we have is, you know, people have been looking at and predicting this post quantum compute era that's always on the verge of happening. So, Raluca, since this technique depends on encryption, is it still viable? What's the risk of that post quantum world changing this in the next couple of years?

[00:26:24] **Raluca Ada Po**pa Absolutely. I definitely believe that, you know, we'll have very performant quantum computers in the near future, even though right now they're still limited in what they can do. The good news with encryption is that we have classes of algorithms that are quantum secure. So, let's take a look at confidential computing and in hardware enclaves. I would say there's two major types of encryptions that they use. One is the encryption during processing, and then there's the encryption for communication and remote attestation for the encryption. During processing, they primarily use symmetric key based encryption algorithms, which are currently known to not, you know, people believe they're not vulnerable to quantum computers. So, in that sense, the main processing of the enclave is not going to be vulnerable to quantum attacks. Again, from what we know so far. But then for the communication and remote testing piece, it relies on, you know, widely used libraries like TLC and TLS has a number of algorithms. You know, like if you have a minor issue that are susceptible to a quantum attacker. Now, the nice thing for that as well is that there are alternatives to class that are post quantum secure, namely, they are not vulnerable to a quantum attacker. So, it really just takes switching to those libraries whenever we feel ready or whenever. You know, in our case, if a customer asks for them, it's not that difficult to switch to them.

There will be some performance degradation. But I would say in the context of enclaves, it's probably very reasonable because most of the work is anyways the computation, and that one I mentioned is already post quantum secure from, from the start. So yeah, I mean it's really a question of when do people want to switch to those algorithms. And you know, I tell people that, hey, even though right now we don't have the quantum attacker, well, if somebody logs your communications right now, they could potentially decrypt them in the future. So, if you worry about what you're talking about right now in exchanging now being visible in a five or so years in the future, then maybe you should start switching to post quantum secure algorithms. If not, you can do it later whenever you start worrying about it.

[00:28:40] **Teresa Tung** I didn't worry about that until you shared that with me. So. But regardless, we can use this technique to even be future proof, right? We can start this and then you have that simplification layer in the software with everything you need to do so that we can start with the use cases.

[00:28:57] **Raluca Ada Popa** Absolutely.

[00:28:58] **Teresa Tung** I wanted to conclude. Right. So, any final thoughts related to that? Things that you want to just make sure this audience really understands.

[00:29:06] **Raluca Ada Popa** I would say that I'm sure that most of the people listening to the podcast have confidential data. I worry about protecting it and also want it to be used in AI. And probably sometimes they feel daunted about bringing these two together. I would say my first, you know, devices start small. Like starting with a well-defined use case that you can wrap your head around and try out confidential computing for that one. See how it works to build trust, and then you can go on from there.

[00:29:37] **Teresa Tung** Wonderful job. Anything for you to add?

[00:29:40] **Rishabh Poddar** I completely agree with your look. And the only other thing that I would add is the future is confidential, right? Like two decades ago. We want encrypting data addressed as an industry practice, but not with simply quotas. One decade ago, we were not encrypting data in transit, but now it's standard everywhere. It's unthinkable that you wouldn't encrypted address and in transit, really the third leg of the stool and the future is encrypting data in use. Confidential computing is I don't know if it's board of me to say, but the only practical way right now to be able to encrypt data and use for large scale workloads. So, it will happen. It needs to happen. The sooner you adopt it, the more future proof you're going to be.

[00:30:24] **Teresa Tung** And yeah, no, I agree.

[00:30:26] **Raluca Ada Popa** And to add to what Rishabh is saying, we actually are running the premier conference in confidential computing. You can find all the materials for it. That confidential computing summit.com, which is great because you know basically the major vendors of hardware out there telling you how they're building the future hardware. The building has confidential protection. The major cloud software vendor is, you know, like Opaque, and we're not showing you how you can secure workloads. Basically, you really get this sense of what Opaque is saying from all those talks that, you know, the future is confidential. And so, yeah, and we made those available, all that material and all the stocks available online for everybody to use as a knowledge base.

[00:31:08] **Teresa Tung** Wonderful. So, thank you both for joining. So, I think there's a lot of good news, right? So, there's a little bit of scary things that generative AI's going to make some attacks possible that was harder to do before generative. AI is going to need us to secure data in different ways. But the good news is that there's a solution, right, with not just confidential computing, but this turnkey software capability that makes it easier for us to apply as an industry. So, we can really the sooner we adopt, it means that we can use data at scale, we can safely share data across our business and use it for generative AI. We can work with partners in different ways. So, thank you so much for joining us.