# GOOGLE'S Royal Hansen

## On security, digital hygiene & the cobbler's children

IN SEEKING AN INTERVIEW WITH GOOGLE'S VICE President of Engineering, Brunswick's Chelsea Magnant held an advantage: She used to work hand in hand with Royal Hansen's team at Google and frequently staffed Hansen in his meetings with DC policymakers. In fact, this article is based on a video interview that began with all the cheer of a reunion between friends.

Yet the guidance Hansen offers below isn't just between friends. He begins from the standpoint of his history in cybersecurity, but broadens into more general thinking about AI, the digital ecosystem and the safety of networks. His thoughts, of greatest interest to cybersecurity experts, warrant the attention of anyone with a keyboard. His goal is to put "into the water," for use by everyone, security practices and information that can help keep digital thieves at bay.

A graduate of Yale with a degree in computer science, Hansen has served as a Managing Director at Goldman Sachs, an Executive Vice President at American Express and a Senior Vice President at Fidelity Investments. In those and other roles, his concern was digital integrity.

**Could you tell us about the road that led you to Google and your role today?**
I started professionally as a software developer, having studied computer science. Yet I'd also studied Arabic and the Middle East, and I always had a deep interest in world affairs and events. So for me it was a kind of natural evolution from software development to software security or the cybersecurity space.

There was a Robert Redford movie called *Sneakers* (1992), where they tried to break into a bank on behalf of the bank, to test their controls. We tried at a little company called @stake to sell the services of

ILLUSTRATION: LINCOLN AGNEW

basement hackers to break into banks as they were building new websites, transferring money, et cetera.

That was my introduction to security. That thread led to working in banking for most of my career, because bankers were the people who cared. The first place that cybersecurity mattered was the banks.

Going on seven years ago I joined Google. It wasn't just the banks anymore. Every industry, every sector, all kinds of technology, whether it was mobile, whether it was AI, whether it was banking or healthcare, whether it was undersea cables or data centers, Google was a place where I could contribute to security that would get into the water. Everybody could benefit from it if we did it well.

**What are the biggest security risks today, particularly given the evolution of Generative AI and just AI more broadly?**

At Google, we've been using AI in our products for 15 years. We had a team in security doing AI kind of in its current state in 2011, 2012. Last year, as you know, two of our scientists won the Nobel Prize in Chemistry for their AI-related work.

AI got baked into Gmail in terms of defending against spam and phishing. It got baked into the Play Store to detect malware in applications. It got baked into the browsers so that the browsers defended against malicious sites. To me, the beauty is we knew how to use AI.

What most worries me is that, throughout industry, cybersecurity professionals might not use the technology to defend. We'll be busy being worried about all kinds of newfangled AI-related threats—which, to be sure, must be addressed. But if we don't use AI for productivity, for insights, for sharing and for automation to make defenses faster, better, we'll fall behind, we'll fail to anticipate the attackers.

The cobbler's children having no shoes—that's my analogy. Let's not be a bunch of security people who are busy worried about artificial general intelligence yet not using it to do our jobs more efficiently and effectively. That, I think, is the first big risk.

The second involves the boom around AI. In that rush—and you've seen this with a few of the other providers of AI or models—people are not doing the basics well because they're so busy worrying about building out a new AI stack or a new chatbot or whatever it is. We can't throw out 25 years of foundational security in the infrastructure, in the data center, in the software, in the databases. This is the second risk: In the rush to follow the shiny new object, we forget about all the hygiene and maintenance and quality that needs to be baked into these stacks.

DeepMind

**The power of AI was on display recently for its role in medicine. Two Google DeepMind researchers in the UK won the Nobel Prize for Chemistry in 2024 for creating an AI-driven model, AlphaFold, that can accurately predict protein structures. Demis Hassabis and John Jumper shared the prize with David Baker, at the University of Washington. AlphaFold has been cited in scientific papers more than 20,000 times. The Nobel Prize notes the enormous potential for drug discovery from their work.**

DEMIS HASSABIS
DeepMind Co-founder and CEO

JOHN M. JUMPER
DeepMind Director

PHOTOGRAPHS: DAN KITWOOD/GETTY IMAGES

In navigating those societal or security or safety questions with policymakers, the private sector is currently playing the leading role. This leads to my third concern: We need the public sector to lead and be willing to think about and accept some tradeoffs. If we don't have a good dialogue on policies and tradeoffs, we will underachieve as a society. Literally there will be fewer people that get new medications for chronic illnesses. There will be fewer innovations in power generation and fusion.

There are going to be enormous benefits in health, in science, in transportation. Think about any sector, and there will be benefits. Just as there was with the Web, however, there will be tradeoffs. Let's not shut down the innovations because we're afraid of the tradeoffs. To me that's just risk management. Security plays an important role in discussing and managing those tradeoffs.

**Earlier you mentioned using AI to protect products and services. To the extent that you can lift the hood and let us have a look, how are you thinking about protecting the models themselves?**

Google's Secure AI Framework (SAIF) became the basis for an industry consortium called CoSAI. It's an industry group working on standards around the idea of secure by design, secure by default. The idea is that the components of all these AI-enabled processes or systems will be baked with security so that the user isn't responsible for it.

That means a lot of work we've done on open-source security to make sure that, as people are consuming these packages of software, they know where it comes from, what's in it, whether it has vulnerabilities and how to use it correctly in their own systems.

SAIF does the same thing but for AI. Think of things like model and data cards, which tell you what's in the model, what data was used, what are the risks, what sort of benchmarking it has against safety or security benchmarks. The first thing is just to make sure we bake security into our AI infrastructure.

Much of it's the same, but there are new elements we're baking it in. Whether those are the filters that help look for prompt injections or poisoning of data, that's all part of SAIF.

Last thing I'd mention is red teaming. Back to the Robert Redford example: For years we have run a very skilled red team internally to test our controls. And we're doing that now more holistically, not just security, privacy, safety. We're doing it in different domains: Red teaming for healthcare is very different than red teaming for financial services. So this is cross-functional red teaming, including building little agentic AI models that auto red team the other models. Once again: Use the models to be better and better at red teaming against the new versions of our models that are used for general purpose.

**How are you using AI for other Google operations?**

I've talked a bit about how it's been baked in in the definition of narrow AI. If you think of narrow AI being like Gmail phishing or spam filtering, there's a lot of that scattered across the Google properties. But the way I think of it now with Generative AI is in four tiers.

First is what I just talked about with SAIF and CoSAI. Make sure the foundations are secure and safe, so we've got teams very active in developing that and deploying it. It's not just a standard. It's what we do internally.

The second is, we don't want to be the cobbler's children, so our teams are just literally using it to summarize emails, to summarize incidents, to automate the generation of a ticket, stuff that's relatively basic but saves a lot of time. Make sure we're first-class users of the more general-purpose capabilities that come out of an AI model.

The third is where we start to use the specific applications, and I'll take coding as an example to extend the more general purpose functions into security. Rather than thinking of security as different from coding, just think of it as secure coding. One example is Gemini (a Google GenAI chatbot), where we're busy working with the teams developing the coding agent such that it does secure coding.

The fourth is the frontier. Finding vulnerabilities, "zero days," as the industry would talk about them, finding vulnerabilities that no one's seen before. We've long deployed very high-end skilled hackers in specific technologies. Now we're using AI to extend their capabilities to find vulnerabilities before anyone else does. Find and fix.

Like Google's AlphaFold found the protein physics (a reference to that Nobel-winning research), we find ways to defend against ransomware using AI.

**How are you thinking about using AI to advance cybersecurity across the digital ecosystem?**

First, we're being transparent. We had a threat intelligence report, which showed the way some of these nation-state groups or other criminal groups are using Gemini or Generative AI generally, although for the moment they're not using it for any novel form of attacks. We were transparent with that report and we need others to be transparent with the attacks they're seeing. We are creating a community of threat-intelligence sharing.

"IF YOU'RE JUST WAITING TO READ ABOUT IT, AND ONE OF YOUR CUSTOMERS IS USING IT, YOU'RE INEVITABLY GOING TO BE BEHIND."

**CHELSEA MAGNANT** is a Director in Brunswick's Washington, DC office and leads the firm's AI Client Impact Unit. She previously worked with Google on tech policy strategy. She began her career with the CIA helping US senior policymakers navigate complex geopolitical issues.

Second, think through sectors. The controls needed to protect those risks in healthcare will be very different than they are in transportation. There may be different tolerances for risk. That is not going to happen because a software developer working on an AI model knows those specific nuances. Each industry is going to have to be great at protecting against bad outcomes, not protecting against AI per se, but AI as part of a process in automation. Anytime you automate something there are risks. This is technology anybody can use—doctors, air traffic controllers. To be clear, everybody ought to use this. Doing security work at that level takes context that the developer doesn't always have.

Third, be agile. The technology is changing not even month to month but week to week. We need to keep updating and iterating. This will be a very dynamic space for many years, so the security people, the risk management people, need to stay current. Again, that's the other side of being conversant and fluent in using the tech. You're also better at anticipating and knowing where things are going.

If you're just waiting to read about it, and one of your customers is using it, you're inevitably going to be behind. Use it in your job. Anticipate the way your users are going to use it in their sector. You can end up with a world where the defense is largely done by agents informed by experts, rather than experts using software to protect against agents. The agents will actually be the defense as well, and we're starting to see that already.

**What advice would you give to ordinary people who are concerned about digital hygiene but lack the resources of a Google?**

Two things. One, use AI. I think you're not going to be a good decision-maker without an intuitive appreciation for what AI is. And the beauty of this tech is it's very easy for anyone to use. My son said everybody's got these things open at college. That's what they live on. So everybody needs to be like that. Everyone needs to be more that way.

Second, be careful whom you pick to work with, because this is a little bit of a gold rush moment. There are people who are going to be trying to make money on the boom. They may not be doing the basics. Be careful that you're not baking in a fly-by-night capability. Avoiding that mistake will be easier if people are fluent in working with the technology. ◆