

Cybersecurity Futurism for Beginners



"Prediction is very difficult, especially if it's about the future!" – Niels Bohr, Physicist, Nobel Prize Laureate

A frequent topic I am asked about is what the future of the SOC looks like. At first glance, this seems like a simple question – but scratch beneath the surface and it's actually really complex. Cybersecurity does not exist in its own little pocket universe. Instead, what happens in security operations is driven mainly by external factors – the economy and macro-economics, internal, external and geopolitics, social and cultural trends, fashions, and of course, natural disasters like pandemics. There is no future of the SOC independent of the Future. So if you want to talk about the future of security operations, you really have to make a whole series of predictions and assumptions about the world in general. That's what makes futurism and trend analysis so difficult. Good futurology synthesises trends from all and any relevant domains and fields into a coherent whole.

It doesn't help that humans are incredibly bad at predicting the future, especially when trying to understand complex issues. We can even empirically study how bad we are at predicting, with definitions for a range of fun and scary cognitive fallacies, like the optimism or recency biases, not to mention the Dunning-Kruger effect. Worst of all – even being aware of these pitfalls doesn't necessarily help prevent succumbing to them.

Case in point, how will Artificial Intelligence develop in the near term, and how will this impact us as security planners and practitioners?

As I write this, Microsoft just announced their Cybersecurity Copilot, a personal AI assistant for security analysts, with industry analysts cautiously declaring that “AI Finally Does More Than Enhance Detection”.

The FT put out an article (paywall) saying that generative AI will affect over 300M jobs in developing economies. A collection of over 1,100 scientists, tech experts, and business leaders including Elon Musk, are publicly calling for a global 6-month pause to consider the “profound risks of AI to society and humanity”

Other commentators like Corey Doctorow, the blogger, journalist, and science fiction author, are already calling generative AI the new “crypto-bubble”.

And by the time you read this, I am sure further developments will have occurred, to drive the conversation forward.

Biases and Black Swans

The examples above represent just a small sample of the viewpoints currently being expressed in the public debate around generative AI. What is remarkable is how widely they diverge in their assumptions and conclusions. We can roughly group them into three vastly different predictions for the

future of generative AI.

Prediction #1: Microsoft and Forrester – Generative AI is production-ready and the impact will be evolutionary rather than revolutionary, at least when applying ChatGPT to incident response.

Prediction #2: Financial Times, Elon Musk, and Co – Generative AI progress is on the slope of massive acceleration that will disrupt work and life for millions of people.

Prediction #3: Corey Doctorow – Generative AI is overhyped and will only have marginal impact, if any.

How can three groups of people, all with highly relevant and applicable backgrounds, knowledge and experience in technology, come to such widely divergent conclusions? More importantly, if they can't agree, how can we decide which future is more likely?

Of course, we can try and assign probabilities to each assertion, maybe based on considerations such as how well informed or knowledgeable the individuals or organizations making these predictions are, or how successful they have been at predicting events in the past. But without truly objective criteria and measures, we will likely just follow our own set of biases.

Most people will fall back on the recency bias – “last year AI didn't work, so next year AI won't work.” As a method this will actually seem to work out well for a while. Until it does not. As Mike Tyson put it, “Everyone has a plan until they get punched in the mouth.” To use a recent real-world example familiar to everyone, when the pandemic hit, many organizations had a multi-year cloud security and remote working transformation program that suddenly had to be ready now.

So far, we've stuck to the easy stuff – the known-knowns and

known-unknowns. These are theoretically in the realm of being predictable – but then there are what Nicolas Taleb calls the unknown-unknowns, which despite all FUD-based cybersecurity marketing to the contrary, are by definition unknowable and unpredictable. These fall into the same bucket as stochastic processes, where randomness and complexity mean that while you can measure the current state of a system, you can't predict the next state.

But as we're learning, there is a deeper body of knowledge and theory around forecasting, trend analysis, futurism, futurology, horizon scanning, or however you want to call it. Much of it revolves around the limits of knowing and knowledge, and approaches to deal with uncertainty.

Futures, not Future

However hard it is to try and predict the future, we still have to try, or we can't plan. But plan we must, whether you're leading a country, a company, a charity or a commune. That's the real reason why futurism exists as a field of study and practice. To make planning for the future less guesswork, futurists and other strategic planners have developed a set of tools to help model future developments, and more importantly, to derive something actionable from them. We're going to be working with two of these in particular:

Horizon scanning is a systematic process used to identify and analyze emerging trends, opportunities, threats, and potential disruptions that could affect an organization, industry, or society in the future. The main goal of horizon scanning is to supply early warning signs of change and to inform strategic planning and decision-making, allowing organizations to prepare for and adapt to future developments proactively.

Scenario planning or scenario method is a strategic planning technique used to explore and prepare for possible future events and uncertainties. It involves developing multiple

plausible future scenarios based on key driving forces and trends, and then analysing the potential impacts of these scenarios on an organization, industry, or system. The main goal of the scenario method is to help decision-makers better understand the risks and opportunities associated with different future outcomes and make more informed strategic choices.

Scanning the horizon for scenarios to analyze

The scenarios we are going to analyze will be based on near-term horizon scanning. Rather than trying to predict a single certain future, we are going to model a variety of alternative future scenarios, all with the same initial starting point – Now. None of our futures are likely to play out exactly the way we anticipate them, but combined across all of them, we should get at least a few things right. This will allow us to be better prepared, and make informed decisions.

AI on the Horizon

We're going to imagine a variety of different futures throughout this series, but all of them will be extrapolated from current trends, with different uncertain assumptions defining the scenarios.

With all of the hype and hustle on the topic of generative AI, we're going to look at some future AI scenarios first.

We can now go back to our original three predictions on how generative AI will impact society in the near term, and also infer a few further ones:

Dead End AI

- AI ends up another hype like crypto, NFT's and the Metaverse.
- AI is overhyped and the resulting disappointment leads to defunding and a new AI winter.

Slow AI

- AI plateaus and stays on our current level
- AI iteratively improves on an evolutionary, not revolutionary path

Controlled AI

- AI progresses rapidly, but is globally rigidly controlled and regulated
- AI progresses rapidly, but is only available to a few great powers

Runaway AI

- A new AI revolution, as disruptive as the agricultural or industrial revolution
- Endgame, Imminent Artificial general Intelligence and the Technological Singularity

We will be imagining all of these scenarios over the next months in a series of articles. For today though, we will quickly discuss the Endgame Scenario – Artificial General Intelligence (AGI) resulting in the Technological Singularity (Kurzweil). We can't make any predictions about it. By definition, and the reason that the term singularity was chosen in the first place, is that its impact will be so momentous that it will cause *unforeseeable* changes to human civilization. We hit a prediction event horizon.

What is remarkable to note is that we've hit a point where reliable prediction becomes exceedingly difficult beyond even short time horizons way before full-blown AGI. It may be the case that the requirement for machine intelligence to be massively disruptive may be far more modest than we thought.

It is worth noting that AI is just one of the many scenario factors that will be relevant for the future of security operations. There are many other interesting and plausible scenario, like what would security look like in a future where

low power consumption, durability and repairability are major factors? What about a world where energy becomes freely available? These and many other factors converge with AI, even acting as inhibitors and catalysts. Some futures are also bright. Let's explore those too.

Cybersecurity Futurism for Beginners