

---

**Today, no matter how wealthy you are, you simply cannot buy a more powerful smartphone than is available to billions of people. In the next decade, access to Artificial Capable Intelligence (ACI) will follow the same trend.**

*The Coming Wave, Mustafa Suleyman [LINK](#)*

---

Over the new year, the pace of generative AI (GenAI) announcements accelerated. HPE's intended acquisition of Juniper for US\$14B is another sign of the shifting infrastructure market, particularly in the network.

On the announcement, HPE's CEO Antonio Nori made the following comment – can you spot the red flag?

*".. when I integrated the HPE networking business into Aruba it was approximately \$1.7 billion and today at the end of 2023 it was \$5.3 billion."*

Jan 2024 [LINK](#)

*Hint: HPE has over 62,000 employees.*



CEOs often provide valuable insights such as this. We delve into the leadership styles of the former Optus CEO and the current Microsoft CEO, each providing important, free lessons.

We also explore recent, selected GenAI announcements and cybersecurity data.

*Previous Newsletters, including this one, are available on our site in pdf [HERE](#)*

**CONTENTS:**

- GenAI:** [Significant announcements in last two months](#)
- [LLMs in your pocket, and on your desk](#)
  - [Platforms](#)
  - [Text-to-Image, Text-to-Video generation](#)
  - [Development](#)
  - [Partnerships](#)
- Leadership:** [Optus & Microsoft CEOs leadership education](#)
- Cybersecurity:** [Serious \(but laughable\) punishment for cyber crime](#)
- [Staggering costs of cyber attacks](#)
- Book:** [The Coming Wave, Mustafa Suleyman, Sept 2023](#)

## GenAI

**“Every single one of our businesses inside of Amazon, every single one, has multiple generative AI initiatives going on right now”**

Andy Jassy, Amazon CEO, 4<sup>th</sup> August 2023. [LINK](#)

The slew of GenAI announcements and their rapid advancements are stunning. Here is a look at selected announcements by function for 2024:

## GenAI – LLMs in your pocket, on your desk

Until recently, the incredible performance requirements of Large Language Models (LLMs) have made it impossible to run them outside of anything but centralised, large scale data centres. In 2024, this will change as AI moves to the edge. Some key announcements include:

### Samsung – AI smartphone and TV [LINK](#)

Samsung introduced the Galaxy S24 series, boasting on-device GenAI tools. This smartphone series enhances communication with features like real-time translations and conversational tone adjustments. The Neo QLED 8K QN900D TV also includes built-in AI capabilities.



### Apple – AI iPhones – Ferret and Flash-Memory LLMs [LINK](#)

Apple launched Ferret, its first open-source Multimodal Language Model (MLLM), reportedly surpassing GPT-4 in image analysis. While the specifics of the AI-powered products for iPhone users remain under wraps, expectations are high for an AI tool that could enhance Siri’s functionality.



While still early in its journey, Apple aims to make Ferret compatible with smartphones. Considering that OpenAI’s GPT-4 has over a trillion parameters, and smartphones currently handle LLMs with around 10 billion parameters, this is no small feat. Basically, Apple’s entire AI value proposition is achieving what is thought as the holy grail of GenAI model serving: running huge LLMs on smartphones. [LINK](#)

**“In layperson’s terms, what Apple has managed is to create the first LLM that can be run efficiently while being stored in flash memory”**

Ferret and flash-memory LLMs mark a pivotal moment in the tech industry as they turn the wildly popular iPhone into new digital products and capabilities.

### NVIDIA – AI Desktops

Nvidia’s announcements of *GeForce RTX™ SUPER Desktop GPUs* and *AI Laptops and RTX-Accelerated AI Software* will bring powerful AI capabilities to a wider range of devices and applications, further expanding the reach of GenAI to the edge.

## GenAI - Platforms

**Google Gemini:** Announced model expansion to more core products, including Google Ads, empowering advertisers with AI-powered capabilities to streamline campaign creation and enhance ad performance.



**Microsoft Copilot:** Enhancements include general availability of Copilot for Microsoft 365.



**AWS Expansion of LLM Choice:** Expansion of its LLM support, particularly through Anthropic's Claude 2.1 model, providing enterprise customers with access to a wider range of AI models for various applications.

## GenAI – Text-to-Image, Text-to-Video generation

Text-to-image improved significantly in 2022. In 2024, AI video generation will be a key trend to watch. It is expected to have major impacts across advertising, movie making and other creative, visual media industries.

For example, take a look at this award-winning video ad for Hotel Chocolat made with the help of GenAI [LINK](#).



**Microsoft: DALL·E 2 in Azure OpenAI Service:** Integration of DALL·E 2, an AI model for generating images from textual descriptions, in Azure OpenAI Service, enabling advanced image generation capabilities.

**ByteDance** (parent of TikTok) introduced video model MagicVideo-V2 in January.

**Google** introduced Lumiere to create realistic videos from text [LINK](#).

**Midjourney** (bootstrapped and likely worth US\$10B) is releasing a text-to-video model in the coming months and quietly disrupting the professional photography industry at the same time [LINK](#).

## GenAI - Development

**Amazon Q Assistant:** Designed to empower employees to obtain answers, solve problems, generate content, and take actions using data and expertise within their organisations.

**NVIDIA TensorRT-LLM Library and AI Workbench:** An open-source tool and workbench that accelerates and optimises inference performance of LLMs on RTX PCs.

**Enhancements to Amazon SageMaker and DataZone:** New inference capabilities for Amazon SageMaker to optimise deployment costs and reduce latency. New GenAI capabilities for Amazon DataZone, simplifying data cataloguing and discovery through automation backed by GenAI.

## GenAI - Partnerships

**Microsoft / Meta:** Expanded their AI partnership with the introduction of Llama 2 on Azure and Windows.

**Runway / Canva:** Runway, which offers AI tools for creatives, also released text-to-video models over the last year to power its AI image & video generation capabilities. The collaboration aims to bring the power of Runway and AI video generation to Canva users. In May 2023, Runway raised US\$141M from Google, Nvidia, and Salesforce Ventures at a US\$1.5B valuation.

**Runway / Getty Images:** A GenAI video model using Getty's library. The model is aimed at Hollywood studios and advertising, media and broadcasting companies to fine-tune their datasets.

**Google / Hugging Face:** This collaboration integrates Hugging Face's platform with Google Cloud's infrastructure, making GenAI more accessible and impactful for developers.

These are only a few selected announcements - there were many more. 2024 will be both exciting and disruptive for GenAI. GenAI will not only change rapidly from within, it will also impact every other industry.

## Leadership lessons from Optus and Microsoft CEOs

Optus and Microsoft CEOs provided some free leadership lessons recently.

In Australia, Optus incurred two catastrophic events within 12 months. The first was a massive ransomware attack, the second a gigantic network failure. In both cases, millions were directly impacted (customers and dollars).

In North America, Microsoft had a (potential) catastrophe over November-December. Having previously invested over US\$10 billion in OpenAI, it made deep commitments to ChatGPT within its products. Microsoft woke up one November morning, to its surprise, to learn that the OpenAI board had ousted its CEO, Sam Altman (pictured). Mass resignations followed.

Optus CEO's response to its first incident was bewilderment, confusion and poor communication to its customers. Having learned little from the first event, the response to the second was worse – no communication until hours later, and only after the minister herself addressed the media. It was a costly, embarrassing disaster.

In contrast, Microsoft's CEO acted quickly and decisively. Soon after the board announcement, he offered Sam Altman a job followed by job offers to all the others.



The result: Optus faces massive damage to its brand and financials. OpenAI changed its mind, decided to keep everyone and the problem was averted.

Microsoft's CEO was nominated as CNN's CEO of the Year 2023. [LINK](#)  
Optus' CEO is no longer there.

## Serious (but laughable) punishment for cyber crime

They found the culprit and wacked him with a feather.

The Australian government recently named Aleksandr Ermakov, a 33-year-old Russian, as the hacker responsible for the Medicare data breach in which he stole data on nearly 10 million customers and cost Medicare an estimated \$50 million.



When the company refused to pay a \$10 million ransom demand, the hackers selectively leaked highly sensitive health records, including those tied to abortions, HIV and alcohol abuse.

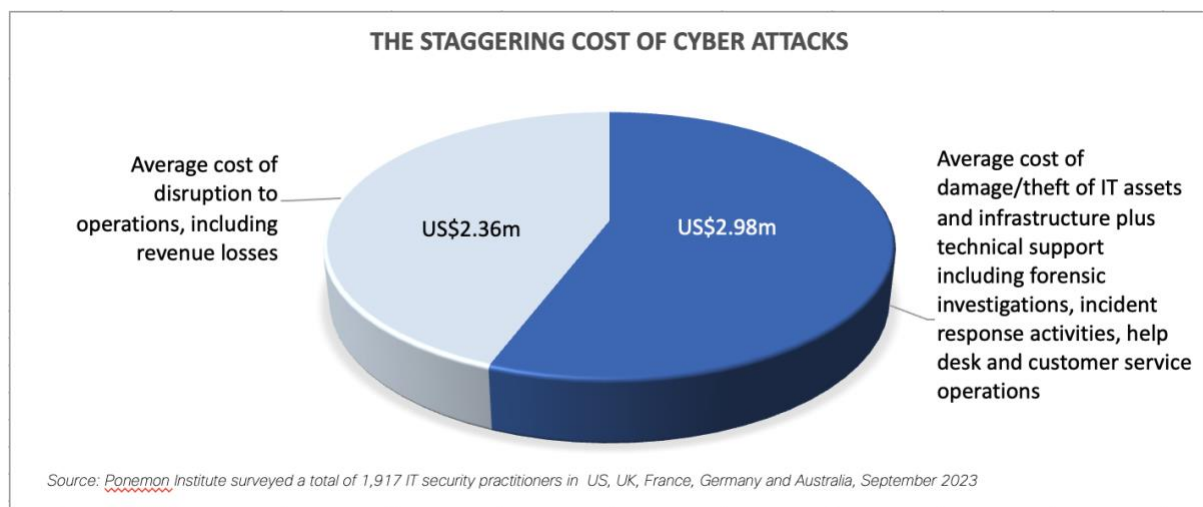
The punishment – “targeted financial sanctions” and a travel ban. So, he can’t visit Bondi and we can’t do business with him, even if you really want to (!)

It is difficult to know what else the government could have done, but it certainly highlights the disparity between attackers and victims. Unfortunately, the reality is that the odds are enormously weighted in favour of the attacker.

One thing is for sure, Aleksandr (pictured) still has a big smile on his face.

## Staggering cost of cyber attacks

A recent Ponemon Institute survey revealed the staggering cost of cyberattacks: US\$5.34 million (the average total annual cost of responding to a compromise).



[LINK](#)

## Book: The Coming Wave

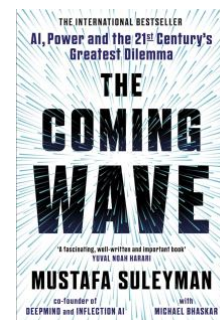
*Technology, Power, and the Twenty-first Century's Greatest Dilemma, Mustafa Suleyman, [LINK](#)*

***“Over the next few years, I believe, AI will become as ubiquitous as the internet itself: just as available, and yet even more consequential”.***

Suleyman is the cofounder of the pioneering AI company DeepMind that was purchased by Google 10 years ago to form the core of its AI capability. He has been at the centre of the AI revolution, poised to become the single greatest accelerant of progress in history.

He sounds the alarm on the unprecedented risks to global order posed by a wave of fast-developing technologies like artificial intelligence and genetic engineering.

He argues that the coming decade will be defined by this wave of powerful, fast-proliferating new technologies. These tools will help address our global challenges and create vast wealth - but also upheaval on a once unimaginable scale.



Suleyman shows how these forces threaten the grand bargain of the nation state, the foundation of global order. We face an existential dilemma - unprecedented harms arising from unchecked openness on one side, the threat of overbearing surveillance on the other. Can we forge a narrow path between catastrophe and dystopia?

Soon we will live surrounded by AIs. They will carry out complex tasks, operating businesses, producing unlimited digital content, running core government services and maintaining infrastructure. This will be a world of DNA printers and quantum computers, engineered pathogens and autonomous weapons, robot assistants and abundant energy. It represents nothing less than a step change in human capability.

His message - we are not prepared.

The book serves as a compelling call to action, urging readers to consider the ethical and societal implications of the coming wave of technological innovation.

Stay connected.

*Kevin*