DXC TECHNOLOGY

Rules of the game: How to secure generative AI within your business

Generative AI may be the new tech buzzword, but its business use cases are still in their infancy. How can organisations make the most of AI without exposing themselves to unforeseen risks?

Rules of the game: How to secure generative AI within your business

According to a **Telsyte survey**, one-third (34%) of Australians are interested in using generative AI chatbots like ChatGPT for tasks such as language translation (44%), brainstorming ideas (34%), summarising long documents (33%), and assisting with research (27%). The study also found that 33% of the Australian workforce interested in the technology plans to use it for work.*

Generative AI may be the new tech buzzword, but its business use cases are still in their infancy. How can organisations make the most of AI without exposing themselves to unforeseen risks?

Artificial intelligence (AI) now affects nearly every aspect of our lives and enterprises. Consider, for example, ChatGPT's tremendous popularity: It was embraced by 1 million users in just 5 days after launch in November 2022, and by 100 million users just 2 months later — and drove the topic of AI into conversations at both the breakfast table and the boardroom table.

ChatGPT is a free and publicly available generative AI chatbot tool developed by OpenAI that is capable of natural conversation with users and assists in general knowledge response. While AI has been a part of technology for many years, generative AI has only recently evolved to be useful to the masses — and the rules are still being created. This is where the discussion becomes interesting.

At its heart, AI recognises patterns and abnormalities in massive datasets — that is why it is such an important component of modern cyber security. AI is already used in the cyber security industry to automate threat detection, respond faster to attacks and help in securing systems and data. The incorporation of generative AI can now boost cyber security professionals' productivity in threat intelligence reporting, security rules generation as well as augmenting incident response process.

We've already begun to see what generative AI can do and what it might be capable of; if the rapid adoption of ChatGPT is any indication, people are eager to play with these capabilities.

Essential issues to consider — now

However, numerous questions have yet to be answered: What rules regulate generative AI? Who uses it, and how? Is the data safe and to be trusted? Where are the dangers? What are the consequences? We are now teaching people to use this form of AI — but what do we know about the responsible way for doing so?

As we do not yet have answers to many of these questions, it is vital that we start the conversation now. Here's a snapshot of what's going on with AI and generative AI in the market.

* "Generative AI Set to Change the Way Australians Live and Work", Telsyte, March 2023: https://www.telsyte.com.au/announcements/2023/3/28/generative-ai-set-to-change-the-way-australians-live-and-work

AI skills are in scarce supply. Australia is not the only country suffering from acute skill shortages, particularly in technology; if shortages already exist in established fields like data science, security and DevOps, it's conceivable that AI professionals will become even more scarce in the future, particularly in organisations' internal IT departments. AI professionals are gravitating towards working with systems integrators (SIs), where they can engage with AI use cases from hundreds of different customer organisations, rather than simply seeing how AI and generative AI are applied narrowly in a single business.

If your company isn't already using AI, it is likely that your employees are. Whether your employees use ChatGPT to help them write emails or code, or for some other purpose, they are probably entering confidential customer and company data into a platform that does not fall under your policies, governance or security — putting company data beyond your control. IT leaders are learning that they have no idea where their data is going, or how it is being used, when employees use public generative AI platforms.

Whether your employees use ChatGPT to help them write emails or code, or for some other purpose, they are probably entering confidential customer and company data into a platform that does not fall under your policies, governance or security — putting company data beyond your control.

We don't fully understand what we're teaching AI, and how it is learning. The value of generative AI lies fundamentally in its ability to learn. Everything we feed into a generative AI platform aids in its learning. However, because we are still in the early stages of development and public AI platforms are essentially "being taught" by tens of millions of users, many unintended outcomes are already apparent. Users, whether purposefully or inadvertently, may perpetuate biases and prejudices. Furthermore, attempts to exploit the system through prompt hacking can result in unintended outputs containing violent, criminal or unlawful content. In short, how people plan to use generative AI models does not always correspond with how the models are actually being used. And we don't know enough about the results to predict them.

Already, we're seeing unfathomable generative AI hallucinations — the generation of erroneous or misleading information that isn't warranted by the AI's training data. While we know that generative AI will usually provide an intelligent response to prompts, the technology may occasionally provide an incorrect response. How can the outputs be trusted if the AI systems — and the humans who use them — are not entirely trustworthy?

Guardrails are still being established. Consider a worst-case scenario: AI essentially makes its own decisions in the context of an organisation's cyber security. Suppose you're using AI's assistance in detecting a cyber risk. But instead of discovering and mitigating the vulnerability, AI decides that the best defence is to attack. Suddenly, you have AI running wild in your company, and you have no control over it. As things stand now, AI may make decisions that we cannot predict.

For many organisations, making AI behave in a "responsible" manner is an afterthought. There are many reasons for this; however, the most common are these two: First, the business is focused more on identifying high-impact use cases for AI than how the AI is being trained, and by whom. This is exacerbated by the second problem: AI solutions are implemented based on existing company policies, rather than the company considering whether these policies are fit for purpose or need to be modified. The result can be adverse simply because the use cases are not well-defined, and the foundation for responsible AI is not there.

Rather than viewing AI as a shiny new tool, business and IT leaders should create strategic frameworks based upon the outcomes they seek, who will be using generative AI, what the security risks are, and what regulations will ensure the responsible and ethical use of AI within the organisation.

Laying the proper groundwork for generative AI. Generative AI is, at its heart, just another computational tool. It's a technology that allows us to make decisions faster and to improve a company's analytics. Rather than viewing AI as a shiny new tool, business and IT leaders should create strategic frameworks based upon the outcomes they seek, who will be using generative AI, what the security risks are, and what regulations will ensure the responsible and ethical use of AI within the organisation.

Begin with no faith, or zero trust. Fundamentally, zero-trust rules should apply to all generative AI- and AI-powered platforms. The basic idea behind zero trust is to recognise that any network, platform or environment, no matter how secure, can be compromised. Just as least privileged access is applied to accounts payable systems or to any other sensitive data, tight regulations and guardrails should be set around who can and cannot be allowed to use AI.

For example, if your employees are permitted to use public generative AI platforms, only a subset should be granted access. Instead of using actual personal or company information, personnel should be educated to use only de-identified or sanitised data. Even the prompts that employees use should be scrutinised for potential disclosure of the company's intellectual property rules, procedures or processes. In brief, allowing access does not mean giving employees unrestricted access to public generative AI platforms; instead, an employee must have a valid reason for each prompt they submit.

Humans must be involved and informed. We must be aware of and prepare for generative AI hallucinations. What if AI generates data that results in an incorrect decision? Can businesses automate without the need for human intervention? For the time being, generative AI can handle the heavy lifting, but the human review of all data and recommendations is required. It's advisable that companies designate specifically approved operators to use — and train — AI platforms, ensuring that only permissible data is entered and that the employees are skilled enough to analyse results and make sound judgements.

Adopt AI-specific policies. To harness the potential of AI, look beyond AI subject matter experts (SMEs) and your technology leaders and build a multidisciplinary AI council that includes SMEs from the legal, ethics and security domains. This AI council should have board-level exposure to ensure that AI governance and ethics are defined at the highest levels of the organisation. Create AI policies and standards that strike a balance between flexibility and safety. You don't want regulations that are too restrictive and discourage innovation, but they need to be strong enough to protect your company, employees and customers/consumers.

How DXC can help

DXC Technology recognises that there will be no one-size-fits-all method to safeguarding the use of generative AI, just as there is no one-size-fits-all approach to security. Every organisation has distinct requirements. Our teams assist our customers in securing their environments by first analysing and mapping threat models and vectors, identifying new technologies they are embracing, and putting in place the essential safeguards to keep the company secure.

This has always been our strength: We don't rely on a single specialist team or area of knowledge to produce a solution. Instead, we begin each engagement with an advisory dialogue to uncover our customers' needs, drawing on capabilities from across our organisation as well as the wisdom and experience of hundreds of use cases to identify patterns and address specific difficulties.

While AI has matured over time, generative AI is still in its early stages in the corporate world. We are all learning to walk together. That said, our team is exposed to what is happening across diverse industries, markets and organisations across the world, and we use this knowledge to help our customers analyse risk and handle all aspects of their security environment swiftly and efficiently.

Let's talk about the future of generative AI in your company and how to safeguard its use. Contact us today to discuss your cyber security and related objectives.

About the author

TM Ching is the Security Centre of Excellence leader at DXC Technology in Asia Pacific, Middle East and Africa. His focus is on improving how the region's Security practice services are built, sold, delivered and managed. He works with service leaders to institute operational improvements and programmes to streamline and optimise how DXC serves its Security customers. TM also works with the Security leadership team to refine its business strategies by identifying emerging trends and pivoting the business quickly to address changing market demands.

Learn more at **dxc.com/au/en/offerings/security** and **dxc.com/generative-ai**

About DXC Technology

DXC Technology (NYSE: DXC) helps global companies run their mission-critical systems and operations while modernizing IT, optimizing data architectures, and ensuring security and scalability across public, private and hybrid clouds. The world's largest companies and public sector organizations trust DXC to deploy services to drive new levels of performance, competitiveness, and customer experience across their IT estates. Learn more about how we deliver excellence for our customers and colleagues at **DXC.com**.