Governing generative AI securely and safely across APAC

By Chris Novak, Senior Director, Cybersecurity Consulting, Verizon

Introduction

Mastering artificial intelligence (AI) is crucial to gain a competitive advantage. Enterprises may unlock a return on investment in a little over a year, with an average return approaching $4 for every $1 invested.[1]

However, achieving mastery hinges on governance or "Responsible AI" instilled as a safeguard and ethical cornerstone. It also means meeting specific threats and vulnerabilities associated with GenAI, including technical aspects, attack vectors, and the need for robust security measures.

Although AI-related attacks currently make up a small percentage of overall attacks in our current 2024 Data Breach Investigations Report (DBIR)[2] , they are still an important topic due to their potential growth in the future.

Deploying new AI use cases can help address some of the world's biggest problems in healthcare, finance, climate change, energy, fire prevention, Industry 4.0, productivity and customer commerce. Verizon, a trusted partner, is deeply committed to helping organisations leverage AI to tackle these critical challenges.

Combining 5G's faster speeds, lower latency, and greater capacity with AI, cloud, and edge computing will enable data to move freely and easily across your business network.[3] However, it's essential to acknowledge that securing this innovation is a profound challenge. It's a challenge that both chief information security officers (CISOs) and the C-suite are grappling with today.

We explore the promise and threat of this transformative technology in the Asia-Pacific (APAC) region, which represents 65% of the world's population[4] and generates over 54% of its gross domestic product (GDP)[5]. Given these stakes, immediate action is required to implement effective governance and security measures to protect these innovations and ensure their ethical deployment.

1 https://news.microsoft.com/source/wp-content/uploads/2023/11/US51315823-IG-ADA.pdf

2 https://www.verizon.com/business/en-au/resources/reports/dbir/

3 https://www.verizon.com/business/resources/articles/s/5g-and-ai-creating-connected-global-business/

4 https://asiapacific.unfpa.org/en/populationtrends#:~:text=The%20Asia%20and%20the%20Pacific,populous%20countries%2C%20China%20and%20India

5 https://www.worldeconomics.com/Thoughts/The-Future-is-Asian.aspx#:~:text=Today%2C%20the%20Asian%20share%20of,for%20less%3A%20about%2033%25

GenAI: The promised land?

Unlike predictive AI, generative AI (genAI) has the potential to create or generate new content, ideas or data patterns that weren't explicitly programmed into our system.

1. Infrastructure enhancement: GenAI enables the processing and transportation of large amounts of data necessary for training more complex AI models, enhancing network performance and reliability.
2. Operational transformation: GenAI impacts internal operations, particularly in sales and engineering. Chat-style genAI tools query past deployments, design choices and customer solutions, democratising access to previously siloed information.
3. Product development and customer service: GenAI offers near real-time data analysis and customer interaction possibilities, such as video stream transcription and instant customer support. This could lead to more dynamic and responsive services.

Verizon Connect recently introduced its advanced AI Dashcam solution to the APAC region. The dashcam acts as a trusty co-pilot for fleet drivers[6]. As you drive down a busy street, the dashcam offers real-time coaching, such as a gentle reminder to keep a safe distance when you get too close to another vehicle.

Globally, enterprises using our 5G platforms are finding novel ways to deal with the rapid digitisation of data from distributed networks. Healthcare organisations are using real-time insights from monitoring devices to improve clinical decision-making.

Supported by AI-enabled solutions like intelligent video surveillance[7] and equipment tracking, healthcare practices are re-imagining ways to help improve diagnostic procedures, operating room analytics and patient safety.

Expanding attack surfaces

However, AI also exposes an attack surface layer that previously hadn't been considered, while driving demand for cloud migration and distributed 5G capabilities.[8]

These expanding attack surfaces, combined with the sophisticated capabilities of generative AI, create a significant risk for enterprises that adopt AI solutions rapidly without full awareness or consideration of what could go wrong.

Alongside these increasingly sophisticated threats, a rudimentary technique continues to drive many modern-day attacks. Vulnerability exploitation remains one of the top three techniques attackers use to gain access to an organisation, with the 2024 Verizon Data Breach Investigation Report (DBIR) highlighting the rapid rise of zero-day vulnerabilities, stressing the critical need for improved patch management and faster response times.[9]

6 https://www.verizon.com/about/news/new-verizon-connect-ai-dashcam-delivers

7 https://www.verizon.com/business/resources/5g/5g-business-use-cases/workforce-productivity/patient-data-analytics/#solution

8 https://semiengineering.com/how-ai-in-edge-computing-drives-5g-and-the-iot/

9 https://www.verizon.com/business/resources/reports/2024-dbir-data-breach-investigations-report.pdf

Putting your emerging AI systems under the microscope may reveal areas for improvement in your organisational AI strategy, which in turn can increase your long-term security posture. This is an important area of focus for IT leaders, with the number of common IT security vulnerabilities and exposures (CVEs) worldwide expected to rise by 25% before the middle of 2025[10].

The dark side of genAI

These approaches, while exciting, can also give rise to privacy considerations. As genAI technologies process and analyse vast amounts of potentially sensitive information, focusing on the accuracy of AI responses and the ethical use of data is paramount.

Large language models (LLMs) often make false claims, called hallucinations. While their answers appear convincing[11], they may only sometimes come from a factually correct source. This raises serious concerns about their reliability and the potential for misinformation in industries like healthcare.

Researchers also discovered they could make ChatGPT, an AI model, reveal its training data[12] just by repeating a word again and again. This unusual request exposed personal details, indicating it's hard to stop AI models from unintentionally sharing sensitive information they've memorised.

Employees who input sensitive data into AI conversational assistants risk unintended disclosures and breaches. This can lead to proprietary information training AI, violating data protection laws and potentially exposing confidential details to unauthorised users or third-party servers.

Businesses need a conscientious approach to AI implementation, balancing innovation with responsibility to protect user data and privacy.

Emerging AI vulnerabilities

The adversarial threat landscape of AI[13] is shaped by analysing real-world cyberattacks and security exercises, revealing vulnerabilities unique to AI systems.

It's an ongoing process, but some dangerous areas are emerging:

- Poisoning the data stream: When attackers manipulate AI training data, introducing errors or malicious triggers. This "poisoning" subtly reprograms the AI, embedding vulnerabilities or backdoors that activate under specific conditions, compromising the system's integrity and reliability.

10 https://www.securitymagazine.com/articles/100426-cves-expected-to-increase-25-in-2024

11 https://www.ox.ac.uk/news/2023-11-20-large-language-models-pose-risk-science-false-answers-says-oxford-study

12 1 Nasr, M., Carlini, N., Hayase, J., Jagielski, M., Cooper, A.F., Ippolito, D., Choquette-Choo, C.A., Wallace, E., Tramèr, F. and Lee, K., "Scalable Extraction of Training Data from (Production) Language Models," arXiv preprint arXiv:2311.17035, Cornell University, 2023. https://arxiv.org/abs/2311.17035

13 https://atlas.mitre.org/

- Evasion through deception: Through a technique known as LLM Prompt Injection[14], attackers craft inputs that deceive AI models, causing misinterpretations and erroneous outputs. This evasion technique bypasses AI defences, exploiting vulnerabilities to execute unintended actions akin to sneaking past security guards.
- Reconnaissance of architectures: Attackers explore AI systems' architectures using a technique known as Discover Machine Learning (ML) Model Ontology[15] to identify weaknesses. By understanding an AI's framework, they pinpoint exploitable vulnerabilities, tailoring precise attacks that undermine the system's defences with surgical accuracy.

Hyper focusing on real threats

The majority (68%) of cyber breaches still involve the human element — including social engineering attacks and errors, excluding malicious privilege misuse, according to the Verizon 2024 DBIR[16].

At many cybersecurity conferences, there's a buzz about outlier examples of AI-fuelled threats, which can sometimes lead to undue alarm.

There has also been concern over deepfake robocalls and the threat of AI being used in new advanced social engineering attacks around elections.

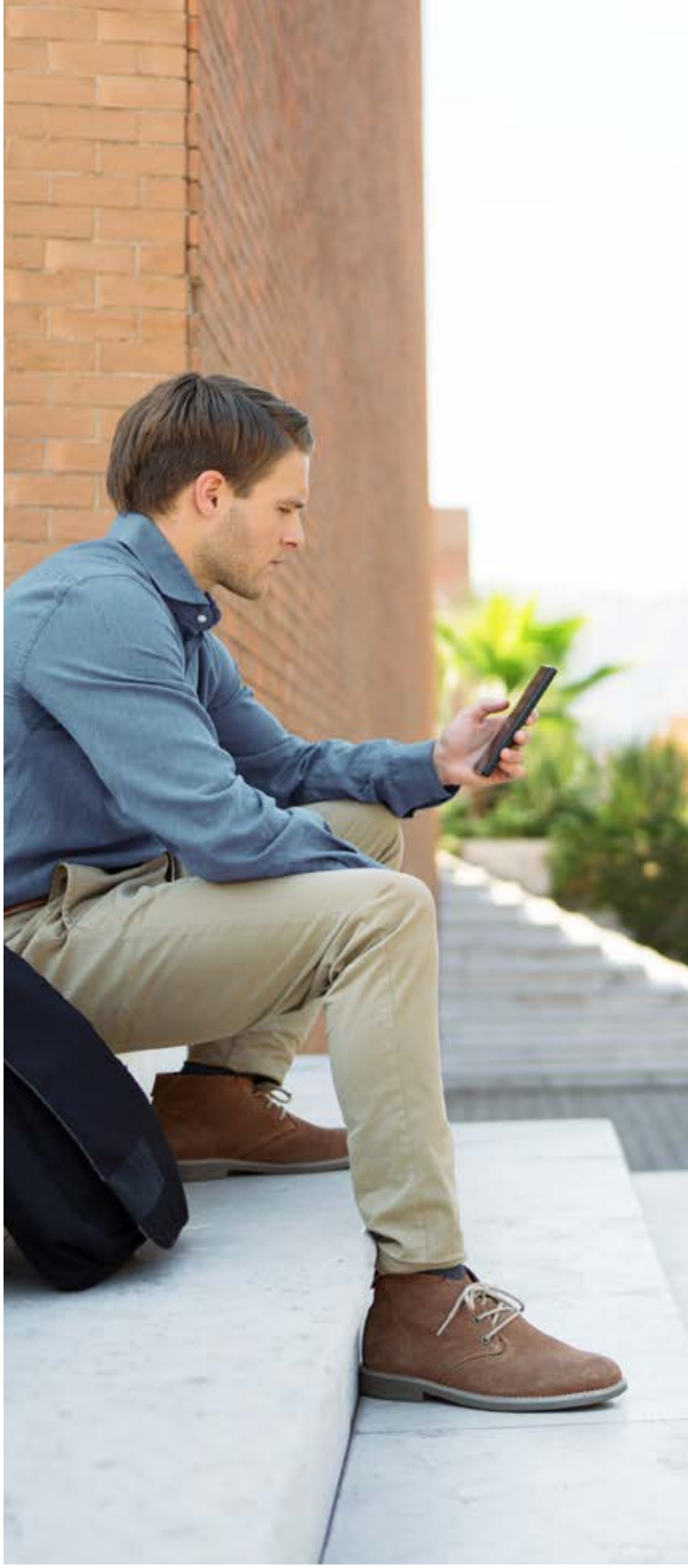
Focus on the actual probability of such attacks:

- The likelihood of widespread attacks using advanced AI techniques is currently very low.
- While there are instances where AI might be leveraged for more sophisticated threats, these remain rare and usually target high-profile individuals rather than the general public.
- Most people are still more susceptible to traditional phishing attacks like email and text messages, rather than AI-powered ones.

A flurry of news around 50 million cyberattacks on an Australian bank[17] would show that only a few are related to full-blown AI-generated sources. This perspective helps us understand the actual risk landscape and focus defences where they are most needed.

The role of AI governance

Businesses are currently playing a cat-and-mouse game with threat actors who will only evolve their techniques and become more sophisticated if forced to.



14 https://atlas.mitre.org/techniques/AML.T0051

15 https://atlas.mitre.org/techniques/AML.T0013

16 verizon.com/dbir

17 https://www.mpamag.com/nz/news/general/bank-chief-reveals-the-volume-of-cyberattacks-banks-are-dealing-with/424684

More concerning risks arise when fragile or incomplete guardrails are considered for emergent AI models to drive innovation or protect companies and governments.

The admission[18] by Australia's Minister for Home Affairs Clare O'Neil that boards tell her they have a list of 30 or 40 people within the government they need to call when they come under cyberattack, illustrates how AI must enhance organisational security and streamline response.

Organisations are excited to build their first internal genAI solution. But, one of the first questions they need to ask is what are they doing to test it? Random tests by those unfamiliar with AI won't reveal if a genAI solution is truly secure. It's like securing a home vs. the Pentagon — the approach must be tailored and quantified.

Placing a regular penetration tester in a complex AI environment invites vulnerabilities, especially as attack surfaces expand into IoT environments and self-optimising plants typical of Industry 4.0. Because when hackers come knocking, they're not playing by any rules. They're out to cause chaos, and testers must be sharp and ready, thinking a step ahead.

Why supply chains are vulnerable

Companies testing AI may also be considered critical infrastructure targets of nation-states with more sophisticated resources. With APAC expected to drive 70% of the world's growth[19] over the next 10 years, securing supply chains is critical.

A recent attack on a prominent Japanese aerospace manufacturer[20] and defence contractor demonstrates "fourth- and fifth-party risk," where the actual depth of supply chain vulnerability extends far beyond the direct partner. With yearly revenues exceeding $1 billion and around 10,000 employees, the company is a significant contributor to broader national defence capabilities. The cyberattack severely disrupted operations, suspending the global website to block further unauthorised access.

Meanwhile, a recent report suggests 83% of Indian companies[21] experienced cyberattacks last year, including supply chain breaches, resulting in significant financial damage. Only 52% of these companies consider themselves prepared for such cybersecurity challenges, underscoring the need for stronger defences in supply chains.

Supply chain risks are a global concern.

It's well known that a cyberattack hits Australian assets roughly every six minutes, including critical infrastructure. The Australian Securities and Investment Commission (ASIC) recently said 44% of companies surveyed had yet to develop a plan to stop data breaches from supply chain partners[22].

Also consider the implications for giant global sporting events such as the approaching Brisbane Olympics. AI attack vectors will likely be the principal driving force behind cyber breaches for these gatherings in 2032, potentially carrying massive financial and infrastructural risk.

During the 2024 Super Bowl LVIII, the Verizon Frontline public safety team worked with dozens of federal agencies to ensure defence teams were ready to combat everything from chemical, biological, radiological and nuclear threats to cybersecurity attacks. The team constantly conducted infrastructure and venue security assessments to stay ahead of potential attacks.

Safeguarding genAI

Minster O'Neil highlights the extreme risks around using and safeguarding AI – not only in Australia but for the entire APAC region: "[We] face the most challenging geostrategic circumstances[23] that we've confronted since the Second World War. We live in a region of strategic competition, and cyber will be integral to how the events of the coming decade play out."

Nations must focus their efforts to get game-ready and become more cyber secure by 2030. To get there, nations in the APAC region, including their private sectors, must start building applications based on 'responsible AI'.

Without strong governance, AI could cause more harm than good, leading to unethical outcomes, biased models and misinformation.

The Verizon 2024 DBIR shows the Public Administration section had the highest number of incidents (12,217) with 1,085 confirmed data disclosures[24].

18 https://minister.homeaffairs.gov.au/ClareONeil/Pages/afr-cyber-summit-18092023.aspx

19 https://www.pwc.com/gx/en/about/pwc-asia-pacific/global-supply-chains-the-race-to-rebalance.html

20 https://westoahu.hawaii.edu/cyber/global-weekly-exec-summary/defense-contractor-japan-aviation-electronics-falls-victim-to-a-cyber-attack/

21 https://timesofindia.indiatimes.com/gadgets-news/over-80-indian-companies-hit-with-cyber-attacks-last-year-report/articleshow/103394017.cms

22 https://asic.gov.au/about-asic/news-centre/find-a-media-release/2023-releases/23-300mr-asic-calls-for-greater-organisational-vigilance-to-combat-cyber-threats/

23 https://minister.homeaffairs.gov.au/ClareONeil/Pages/afr-cyber-summit-18092023.aspx

24 https://www.verizon.com/business/resources/reports/dbir/2024/industries-intro/public-administration-data-breaches/

Mobilising around genAI

AI principles aim to steer technology toward societal benefits in the APAC landscape. Yet, the real test lies in their enforcement across this diverse region.

While ethical guidelines aim to curb AI's misuse without firm regulations[25], the gap between vision and actionable implementation remains critical:

- Adopting guidance and tools to operationalise AI principles varies greatly among APAC nations, raising questions about consistency.
- Legislation efforts, while promising, face the challenge of keeping pace with AI's rapid evolution.
- Ambitious national strategies across countries like Japan, Malaysia and Australia set high expectations for AI's role in development.

Notably, three front-runners have emerged in AI regulation across the APAC region, each carving out a path others might soon follow:

- Singapore is a leader in developing practical tools for AI safety. Recently the Infocomm Media Development Authority (IMDA) published for consultation the proposed Model AI Governance Framework for Generative AI[26]. This will support finalisation of the Model AI Governance Framework in mid-2024.
- South Korea's bold AI Act[27] will cast a wide net if implemented, from general use to high-risk applications, marking a legislative first.
- Mainland China has taken a rules-based approach, with specific AI regulations[28] crafting a comprehensive approach to AI management.

Businesses should create AI councils to fully harness and leverage AI's power and promise. Doing so can help them transform business models, marketing, knowledge management and software engineering responsibly and securely.

Risk management must be fully embedded and integrated to succeed, not playing catch-up. A risk quantification service[29] can help identify potential platform weaknesses and AI compliance gaps.

Verizon cybersecurity assessment includes Red Team Penetration Testing that uses simulated attacks to evaluate threats, including AI. Penetration Testing can run automated tests that probe systems to seek out attack vectors and vulnerabilities, and support target selection.

Additionally, Verizon has implemented AI governance measures, requiring data scientists to register AI models for review and scrutinising large language models (LLMs) to address potential bias and toxic language. These efforts

25 https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/financial-services/deloitte-cn-fsi-acrs-gai-application-and-regulation-in-apac-en-231204.pdf

26 https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2024/public-consult-model-ai-governance-framework-genai#:~:text=SINGAPORE%20%E2%80%93%2016%20JAN%202024&text=The%20AI%20Verify%20Foundation%20(AIVF,last%20updated%20in%2020201

27 https://carnegieendowment.org/research/2024/02/koreas-path-to-digital-leadership-how-seoul-can-lead-on-standards-and-standardization?lang=en

28 https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117

29 https://www.verizon.com/business/products/security/cyber-risk-management/governance-risk-compliance/

align with the broader push for responsible AI and are integrated into their governance, risk management, and compliance (GRC) services.[30]

These tactical approaches need to conform to cybersecurity governance, risk management and compliance (GRC), which are critical pillars in the push for zero-trust security.

All organisations in the APAC region must work together to overcome genAI gaps by 2030. Disparities in AI governance, from ethical guidelines to legislative frameworks, could lead to uneven advancements and vulnerabilities, impacting everything from economic growth to cybersecurity.

Singapore and the U.S. recently linked their AI governance frameworks[31], a groundbreaking move and a significant step toward international alignment on AI policies.

Both countries collaborated with Japan and Australia[32] to guide organisations on how to use AI systems securely. They have also teamed up to craft the first globally agreed-upon Guidelines for Secure AI System Development[33] to enhance the security of AI systems.

The Secure-by-Design framework[34] for AI and tech means even small businesses can start safely without big IT teams.

In total, 21 global agencies are working under the framework to guide AI system developers in making cybersecurity-focused decisions throughout the development lifecycle.

Steering genAI safely into the future with Verizon

While many hope genAI will be "the great equaliser,"[35] the reality is different. Most government agencies and organisations face knowledge and talent gaps that threaten the safety and security of ordinary people across the region.

Businesses should act now, not tomorrow, to quantify their AI risk[36]. The direction and energy for this action must come from the C-suite – security is a culture that must be driven from the top, rather than left as a function of the security team.

Verizon can help cyber teams forge a cross-functional AI steering team, which is a critical step before building an organisation's first GenAI application. Working collaboratively is vital to staying ahead in AI and ensuring cyber safety.

30 https://venturebeat.com/ai/verizon-exec-reveals-responsible-ai-strategy-amid-wild-west-landscape/

31 https://www.mci.gov.sg/media-centre/press-releases/singapore-and-the-us-to-deepen-cooperation-in-ai/

32 https://www.cyber.gov.au/resources-business-and-government/governance-and-user-education/governance/engaging-with-artificial-intelligence

33 https://www.cyber.gov.au/sites/default/files/2024-04/Guidelines%20for%20Secure%20AI%20System%20Development%20%28November%202023%29.pdf

34 https://www.cisa.gov/resources-tools/resources/secure-by-design

35 https://www.weforum.org/agenda/2024/02/generative-ai-society-equalizer/

36 https://www.verizon.com/business/products/security/cyber-risk-management/governance-risk-compliance/cybersecurity-assessments/

Over the last several years, our team has spent significant resources developing specialised or applied AI to solve everyday tasks related to network performance optimisation, identifying trends, generating demand and enhancing customer service.

Verizon trains large datasets to perform finite, well-defined tasks — AI is tailored to address specific operational or business needs. We understand the benefits and the risks.

Verizon network processes 70 billion data points daily, feeding these into advanced AI systems. This data comes from a diverse array of 29,000 sources, showcasing the vast scale and complexity of the digital ecosystem[37].

Our advice is customised for business needs, generating a strong defence plan based on solid data and standards, with detailed security reports and comparisons to industry benchmarks.

Delivering AI-secure threat detection and analysis

Within this framework, we can use AI to refine the approach to bad actors:

- Continuous monitoring: AI systems vigilantly monitor network activity 24/7, uncovering anomalies that might indicate a threat that humans could easily overlook.
- Automated Penetration Testing: These tests simulate cyber- attacks on computer systems, networks, or web applications to identify vulnerabilities that could be exploited.
- Traffic analysis: AI distinguishes between normal and suspicious traffic, enhancing the detection of sophisticated cyber threats.
- Phishing detection: By learning the characteristics of phishing and spam, AI helps pre-emptively block malicious emails.
- Malware identification: AI tools analyse known malware samples to recognise new variants and zero-day (previously unknown) threats.
- Password security: AI can generate and recommend complex passwords that are difficult to crack.
- Task automation: Routine cybersecurity tasks are automated by AI, freeing up specialists to tackle more strategic issues.

GenAI is a powerful solution that can be used to help businesses, their employees and customers. Moving in this direction will not only provide a competitive advantage in the blossoming APAC marketplace – it will also make the region a safer, more secure place for its people.

For more information on how Verizon can help with AI security call.

Australia +61 2 9434 5000

Singapore +65 6248 6600

Japan +81 3 5293 9000

37 https://www.sdxcentral.com/articles/interview/verizons-70-billion-network-data-points-highlight-genai-potential-and-challenges/2024/02/

This handbook was created by InnovationAus.com, with information and input from Verizon.