Threats *to come*

By CHRISTOPHER FAM
lifestyletech@thestar.com.my

MALAYSIA ranked 37th out of 161 countries on cybersecurity firm Surfshark's list of the countries most affected by data breaches in its *Wrapping Up 2025: Global Data Breach Statistics* report.

Last updated on Feb 3, 2026, the report states that 2025 alone saw a whopping 553,722 Malaysian accounts posted to publicly available databases, containing personally identifiable information such as email addresses.

This is reflected in data from CyberSecurity Malaysia's Cyber Security Response Centre (Cyber999), which recorded a total of 5,735 cyber incident reports as of the first nine months of 2025, an increase of 1,076 cases when compared to the same period in 2024.

Fong Choong Fook, founder and CEO of cybersecurity firm LGMS, believes the reported numbers are likely on the conservative end of the scale.

"One of the prime reasons is that, a lot of the time, when cyberattacks and scams happen, not everyone is willing to stand up and say, 'Hey, I've been hacked' or 'I've been breached'. The actual numbers are usually way higher than this," he says.

From the perspective of Dr Megat Zuhairy Megat Tajuddin, the chief executive of the National Cyber Security Agency (Nacsa), the seemingly large number of cases is mainly due to the rapid growth in online users outpacing the adoption of good cyber hygiene practices.

"Many users still click on suspicious links, use weak passwords, download pirated software, or install unknown apps, especially on mobile devices, which can later lead to increased exposure to phishing, malware, spyware, and account breaches.

"Key drivers include low public awareness, unsafe downloads, social engineering where scammers use lies and manipulation to trick you into giving away your details, data leaks originating from third-party sources, and poor organisational security practices.

"Another issue is systems developed without security functions or proper testing built in from the start," he says, adding that this is something that everyday users should be concerned about.

Megat Zuhairy adds that while the figures highlight Malaysia's position in data breach statistics, they "do not necessarily indicate that major system breaches have occurred".

"As a digitally connected nation, Malaysia experiences the same everyday cyber risks as many other countries. Based on these reports, there is no evidence that Malaysia's critical national infrastructure has been systematically compromised," he says.

However, Nacsa anticipates an increase in familiar threats like phishing attempts, mobile malware, spyware, scam campaigns, credential theft, and artificial intelligence (AI)-driven scams this year.

Fong shares similar expectations, noting that what sets 2026 apart is the growing risk posed by new AI technologies such as agentic AI, and the new avenues they create for threat actors.

"Now, with the rise of AI capabilities, many attackers are also relying on AI tools to create more convincing attacks and more convenient attack techniques. Basically, AI has been weaponised to carry out the work for hackers," he says, adding that rising cyber incidents are largely driven by the ability to automate them.

Fong says that this could vary from more convincing phishing emails – now without obvious tells like poor grammar – to impersonating people through deepfakes, making for a potentially serious outcome if they go undetected.

"One thing for sure is that Malaysians should always stay vigilant and aware of what is happening in terms of cybersecurity trends and the latest attacks affecting the world," he says.

New tools, new threats

The World Economic Forum's (WEF) *Global Cybersecurity Outlook 2026* report similarly predicts that such AI tools will be leveraged by threat actors this year, while another WEF report, *AI Agents In Action: Foundations For Evaluation And Governance*, goes into further detail.

It states that without strong governance, AI agents can accumulate excessive privileges, inadvertently propagate errors and vulnerabilities at scale, and even be manipulated through design flaws or prompt injections.

Technology and consulting firm IBM defines prompt injection as a type of cyberattack targeting large language models, in which malicious instructions are hidden within prompts to manipulate outputs, extract sensitive data or spread misinformation.

Coupled with the growing accessibility of agentic AI, thanks to popular open-source offerings such as OpenClaw, this opens up a whole new set of cybersecurity challenges to navigate.

> **Emerging technologies, including AI tools and automated systems, can pose new risks if not configured securely.**
>
> Megat Zuhairy Megat Tajuddin

Megat Zuhairy notes that self-hosted agentic AI systems like OpenClaw, where users run them on their own devices instead of relying on cloud services, can introduce vulnerabilities for users.

"Emerging technologies, including AI tools and automated systems, can pose new risks if not configured securely.

"If they connect to files, browsers, email, or internal systems, they may be exploited as an additional attack surface for data leakage, unauthorised actions, or malware delivery," he says.

Know the risks

Dr Manjeevan Singh Seera, an associate professor at the Monash University Malaysia School of Business, warns that giving AI agents this level of system access carries significant security risks.

"Unlike normal chatbots, these systems are not just giving answers but actively taking actions on your behalf using your files, accounts and credentials. When you give an AI that level of access, it is almost like leaving the door to your digital house open.

"The biggest risks include accidental deletion of data, unauthorised transactions, exposure of private information, and the possibility that hackers could hijack the agent to act using your own credentials," he says.

He highlights that if subject to a prompt injection attack, an AI agent could stand to do more damage than it would otherwise.

"In a traditional chat interface, a manipulated prompt might only affect the text generated. However, in an autonomous agent, a hidden instruction could cause the system to delete files, send emails, transfer data or change settings.

"Because these agents can read external content like emails or websites, a malicious instruction hidden inside that content could trigger harmful actions without the user realising it," he says.

This effectively means that if a self-run AI agent is compromised while having ongoing access to a user's system and payment details, it could potentially take action without requiring approval.

"That means if it misunderstands instructions, malfunctions, or is compromised by an