**An AI-infused world needs matching cybersecurity**

Last year, an incident of a frantic mother who had received an ominous call from “kidnappers” who had ‘kidnapped’ her daughter, raised an alarm in the U.S. Senate about the detrimental impact of artificial intelligence. The news took the nation by a storm as the said “kidnappers” and the daughter’s voice were nothing but hackers utilising generative AI to extort money. With such instances on the rise, the human perception of what is real and what is merely generative AI is slowly eroding.

**Sophisticated cyber threats**

While it is true that generative AI has exceptionally transformed how we operate, with its integration into sectors such as education, banking, health care, and manufacturing, it has also transformed the paradigm of cyber-risks and safety as we know it. With the generative AI industry projected to increase global GDP by as much as $7 to $10 trillion, the development of generative AI solutions (such as ChatGPT in November 2022) has spurred a vicious cycle of advantages and disadvantages. According to a recently published report, there has been a 1,265% increase in phishing incidents/emails, along with a 967% increase in credential phishing since the fourth quarter of 2022 arising from the exacerbated utilisation/manipulation of generative AI.

With sophisticated cyber threats on the rise, organisations and individuals are susceptible to the novel avenues of cyber-attacks, pushing firms to adapt to ever-evolving technology. As per a study conducted by Deep Instinct, around 75% of professionals witnessed an upsurge in cyberattacks in the past year alone, while 85% of the surveyed respondents have attributed the increased risk to generative AI.

It becomes imperative now, more than ever, to develop solutions through collaborative avenues to safeguard confidential information, identities, and even human rights.

As generative AI continues to mature, newer, more complex threats have arisen: through cognitive behavioural manipulation, critically dangerous incidents have surfaced, with voice-activation toys and gadgets that encourage dangerous behaviours in children and/or posing a grave threat to one’s privacy and security. Simultaneously, remote and real-time biometric identification systems (such as facial recognition) have further jeopardised the right to privacy and massively endangered individuals on several occasions in recent times.

While generative AI has significantly impacted productivity across the industrial realm with 70% of professionals reporting increased productivity, increasing manipulation via generative AI (specifically over the past couple of years) has resulted in the the spiralling vulnerability of organisations to attacks, with most organisations citing undetectable phishing attacks (37%), an increase in the volume of attacks (33%), and growing privacy concerns (39%) as the biggest challenges.

The recent identification, by several cybersecurity conglomerates, of complex hacker groups using generative AI solutions has raised an alarm — with AI models being leveraged for translating and identifying coding errors to maximise the impact of cyberattacks.

With such multifaceted cyberattacks on the rise, robust initiatives have become necessary. While stringent ethical and legislative frameworks are underway to combat growing cybercrimes due to AI, loopholes and a lack of industrial understanding/comprehension in regulating generative AI persist.

The Bletchley Declaration

Considering the growing concerns amidst increasing misuse of generative AI, it becomes imperative to safeguard consumers against the challenges posed by such advanced technologies, allowing them to navigate digital spaces safely.

World leaders, too, have initiated collaborative efforts to understand the potential catastrophic harm caused by the detrimental utilisation of AI, as seen in the recent signing of the Bletchley Declaration at the AI Safety Summit. The countries that signed the agreement include China, the European Union, France, Germany, India, the United Arab Emirates, the United Kingdom and the United States.

At the institutional level, stern policy-led efforts are pivotal to bolstering the stance against increasing challenges via solutions such as enhancing the stance for watermarking to identify AI-generated content. This could aid in reducing cyber threats from AI-generated content, warning consumers to take appropriate actions. Further, a collaborative effort between institutional and industrial stakeholders could necessitate the process of improving and implementing a realistic, practical, and effective framework, with the inclusion of feedback from the public to further strengthen the drafting of these regulations.

**Foster digital awareness**

At the corporate level, greater emphasis is required to accommodate digital awareness via occupational media and digital literacy training sessions, fostering robust digital fluency in the workspace while identifying and tackling gaps in digital knowledge among employees. This could further equip the workforce to efficiently navigate the digital landscape, identify credibility, and verify the sources for authentication.

However, for a truly holistic approach to cybersecurity in an AI-driven world, we cannot overlook the crucial role of non-governmental organisations and other outreach organisations that introduce individuals to the wonders of the digital world, and simultaneously equip them with the essential tools of cyber literacy. By fostering a digitally savvy citizenry from the ground up, we can build a more robust defence against the evolving threats in this AI-driven digital landscape.

As we move towards developing more sophisticated systems and technologies, collaborative efforts are paramount to harbour a sense of security, enabling individuals and organisations to further empower communities to safeguard their personal interests and identities.