



September 2026 Cyber News

On behalf of the Yuval Ne'eman Workshop for Science, Technology, and Security, we are pleased to share with you some of the most notable events and developments that occurred in September 2026.

September 1 – Japan Reestablished Foreign Ministry Cybersecurity Policy Division to Advance Cyber Diplomacy – Japan's Ministry of Foreign Affairs (MOFA)

reinstated its Cyber Security Policy Division, under the National Security Policy Division of the Foreign Policy Bureau. MOFA first established a division under the same name in 2016, before later merging it into the National Security Policy Division. The reinstated division is intended to play an active role in promoting the rule of law in cyberspace and shaping international rules in line with Japan's fundamental principles, while strengthening coordination with allies, like-minded countries, and other partners. Between 10 and 20 officials from the National Security Policy Division will concurrently serve in the new division and advance Japan's cyber diplomacy, including by providing capacity-building assistance to developing countries in the cyber domain.

September 9 – FBI Released New Cyber Strategy – The strategy outlines the approaches of its Cyber Division for countering malicious cyber activity targeting the United States. The strategy is built around four lines of effort. To disrupt threat actors and impose costs on them, the FBI will advance joint, multi-phased operations with government, international, and private-sector partners. To support victims of cyber incidents, the FBI will develop automated mechanisms for sharing cyber threat intelligence, to reduce the time between threat detection and the dissemination of information to partners. The FBI will also deepen private-sector cooperation through programs such as InfraGard, which connects critical infrastructure owners and operators with the FBI for cyber threat information and training. Finally, the FBI plans to strengthen its cyber capabilities through the adoption of AI tools, including agentic AI systems, to enhance its defensive and disruption capabilities. These tools will be used to accelerate

malware analysis, map threat actor infrastructures, and support the attribution of cyber activity.

September 10 – ECSO Examined Europe's Shift toward Offensive Cyber Capabilities

– The European Cyber Security Organisation (ECSO) published a [paper](#) on the development and deployment of offensive cyber capabilities by the European Union and European countries, highlighting key barriers that could impede these efforts. According to the paper, the rise in cyberattacks against European institutions and critical infrastructure, alongside the expansion of hybrid threats, has led Europe to recognize that deterrence in cyberspace cannot rely solely on defensive measures and may also require demonstrating credible offensive capabilities. At the European level, this shift was reflected, for example, in [the European Defence White Paper](#) published in March 2025, which proposed, for the first time, developing a voluntary support framework for offensive cyber capabilities together with EU Member States as a credible deterrent. However, the transition is constrained by a shortage of skilled and experienced offensive cyber experts. At the same time, Europe lacks a framework that clearly defines which offensive cyber capabilities private companies may develop or deploy, under what circumstances, and how responsibility should be divided between the public and private sectors.

September 10 – Anthropic Detailed Malicious Uses of Claude in Cyber Operations –

Anthropic published [a report](#) on the use of its Claude large language model (LLM) by various actors between December 2025 and August 2026, including in cyberattacks. Hackers used AI to varying degrees, from developing malware and phishing kits to conducting largely autonomous attacks through multi-agent AI frameworks. In other cases, hackers directed Claude to carry out attack activities while retaining control over target selection. One example involved the hacker group GTG-20006, which is believed to be linked to the Russian group Midnight Blizzard. The group used AI to manage and modify attack tools, including a credential stealer targeting passwords stored in web browsers, and to monitor their detection by cybersecurity defenders. AI agents modified the tools until their malware was no longer detected. The group targeted government, military, and diplomatic entities in Ukraine and Europe, focusing on suppliers and supply chains involved in providing military drone technologies to Ukraine.

September 14 – China Released Updated AI Safety Governance Framework –

China's National Technical Committee on Cybersecurity Standardization (TC260) published [the third version](#) of its framework for governing AI safety risks, structured around three main components. First, it identifies inherent safety risks in the processes and infrastructure involved in AI development, such as data poisoning attacks, as well as risks arising from AI use, including autonomous cyberattacks conducted by agentic AI. Second, it presents technical measures to address these risks, including red-team testing during model development to identify and mitigate security vulnerabilities. Finally, the framework sets out policy principles for risk management. Among other measures, it calls for AI safety curricula and training programs throughout the education system to develop a skilled AI safety workforce. The document also includes several annexes, including a framework for managing potential risks throughout the agentic AI development and use lifecycle and principles for advancing trustworthy AI, supporting the implementation of China's October 2023 [Global AI Governance Initiative](#).

Make sure you don't miss the latest on cyber research.

[Join our mailing list](#)



