Webinar

Managing Cyber Risks in the Era of Powerful AI Models

KEY FACTS

From the webinar "Claude Mythos and co.: Understanding and Managing New Cyber Risks for Financial Institutions" presented by Bank-Verlag in cooperation with the Association of German Banks (BdB) and the Bankenakademie on 24 June 2026.

The webinar highlighted a range of perspectives, including the supervisory viewpoint, assessments from the national cybersecurity authority and practical insights from several financial institutions.

The situation from Bafin's perspective

Key message: Cyber risks remain a key focus for supervisors due to the high threat level and the potential for cyber incidents to spread quickly beyond individual institutions, affecting both the wider financial system and the broader economy. Digital innovation is seen as essential to maintaining a functional, competitive and resilient financial system. At the same time, increasing utilisation of digital technologies is expanding the cyber risk landscape.

This applies equally to the opportunities and risks associated with new frontier AI models. From a supervisory perspective, full implementation of DORA and strong cyber hygiene provide a solid foundation for addressing the current challenges. However, additional measures are required. Financial institutions must be prepared to assess and remediate a growing number of new vulnerabilities within short timeframes, or to mitigate them using appropriate measures. A risk-based approach is vital, with priority given to protecting critical systems. Organisations must also develop strategies for legacy systems that can no longer be patched. In doing so, they need to consider not only risks within their own environments but also those affecting ICT third-party service providers.

Key drivers: Complex IT environments, geopolitical conflicts and concentrations and dependencies are key factors contributing to the elevated risk landscape. The growing use of AI is creating new risks. Particular attention is currently focused on powerful frontier AI models such as Claude Mythos 5, GPT-5.5 and MDASH. These models are expected to be capable of identifying and exploiting previously unknown vulnerabilities in ICT systems within a short period of time. As the provision of these models is concentrated among a small number of providers, this development could further increase dependencies on individual providers in third countries. The conditions governing access, data protection and legal requirements are currently evolving on an almost daily basis.

Further Reading: The importance of information sharing, discussion and the exchange of practical experience and best practices was highlighted throughout the webinar. Nikolas Speer's full keynote is available here: Aktuelles & Presse - „Claude Mythos & Co.: Neue Cyberrisiken für Finanzunternehmen verstehen und bew ältigen" - Bafin

PARTNER ORGANISATIONS

bankenverband bankenakademie

BVbank-verlag

Frontier AI: accelerating cyber risk

Frontier AI models such as Mythos are increasing cyber risk in both the short and long term. They enable ICT vulnerabilities to be identified and exploited more quickly, lower the level of expertise required for complex attacks and increase the speed, scale and sophistication of cyberattacks.

- **Lower barriers for attackers:** AI reduces the level of expertise required and shortens the time needed to develop and deploy exploits.
- **Greater speed and scale:** AI is expected to lead to an increase in zero-day exploits, software supply chain attacks and attacks on service providers.
- **More security incidents expected:** The emergence of large numbers of vulnerabilities and patches within short periods of time is increasing the pressure on organisations to respond.

See also: Impact of Developments in Artificial Intelligence on the Cybersecurity of Organisations (BSI publication of 22 June 2026)

Patching under time pressure

Attacks are occurring at a pace that exceeds traditional patching cycles. As a result, vulnerability and patch management processes must become significantly faster, scalable and subject to active follow-up.

- **Understanding your IT infrastructure and critical assets:** This enables effective prioritisation.
- **Strengthen vulnerability and patch management:** Identification of vulnerabilities, patching and incident response should be automated or semi-automated.
- **Create transparency:** Vulnerabilities should be visible to technical teams and developers, while KPIs create transparency for management bodies.
- **Actively track progress:** A dedicated vulnerability manager can support the tracking and follow-up of remediation activities.
- **Address key challenges:** These include maintaining business continuity, replacing legacy systems, automated testing and keeping a human in the loop.

Key takeaway: Patch and incident management processes must be scalable, enable effective prioritisation and strengthen follow-up.

Cyber hygiene as a strong foundation

Cyber hygiene, defence in depth, zero trust, security by design, attack surface reduction, security in depth and crisis response were identified as key areas of focus.

- **Implement core cybersecurity measures:** Defence in depth and zero trust approaches should be further strengthened.
- **Apply security by design:** Security requirements should be taken into account from the outset when designing, developing and integrating large language models (LLMs).
- **Improve cyber hygiene:** This includes patching, end-of-life management and penetration testing.
- **Reduce the attack surface:** Organisations should seek to minimise their attack surface, prioritising external assets and assets accessible via the internet.
- **Adapting modern AI technologies:** AI technologies should be used for cyber defence and vulnerability discovery.

Key takeaway: Strong cyber hygiene reduces the attack surface, reinforces core controls and creates the conditions for responding effectively to AI-driven threats.

Ensuring resilience: detect, contain and recover

The Digital Operational Resilience Act (DORA) has been in force since 2025 and provides the foundation for the resilience of financial institutions. A rapidly evolving threat landscape, dependencies on third parties and cyber risks require enhanced and well-rehearsed emergency, business continuity management (BCM) and ICT incident response structures.

- **Anomaly detection:** Technical safeguards include anomaly detection capabilities.
- **Strengthen containment:** Containment capabilities and network segmentation are key elements of resilience.
- **Enhance response drills:** Improved response drills and cyber incident response playbooks strengthen an organisation's ability to respond to threats.
- **Strengthen restoration capabilities:** Restoration capabilities and recovery plans form an integral part of resilience planning.
- **Align exercises with emerging threats:** Crisis exercise scenarios should take frontier AI into account.

Key takeaway: Resilience remains the foundation; containment capabilities, restoration capabilities and response drills are becoming increasingly important.

Keep an eye on third parties and supply chains

Risks come from attacks on third parties, service providers, software supply chains and compromised third-party libraries. Third-party software and open source software also represent a significant area of risk.

- **Third parties:** A higher level of security at third-party providers should be measured and required.
- **Suppliers and vendors:** Software suppliers and vendors should be held accountable when software is commissioned or purchased.
- **Third-party libraries:** A growing number of compromised third-party libraries is increasing risk.
- **Open source and version control:** Specific software versions should be defined rather than relying on 'latest' releases.
- **In-house development:** AI-based vulnerability detection and remediation should be integrated into internal processes.

Key takeaway: Third parties, service providers, supply chains, third-party software and third-party libraries all form part of the current risk landscape.

Legacy and end-of-life systems: key areas of risk

Legacy systems present a challenge for vulnerability and patch management. End-of-life management is a key element of improved cyber hygiene.

- **Take account of legacy systems:** Legacy systems make remediation, patching and automation more difficult.
- **Include end-of-life management:** End-of-life management is a part of improved cyber hygiene.
- **Enable prioritisation:** A thorough understanding of the IT infrastructure and associated assets allows effective prioritisation.
- **Strengthen cyber hygiene:** Patching, end-of-life management and penetration testing are key measures.

Implementation: think big, act with focus

Implementation approaches vary depending on the size and structure of the institution, but the key priorities are largely the same. For example

- ✓ **Large bank:** Typical priorities include establishing a steering board and task force, improving prioritisation and risk reduction, strengthening vulnerability and patch management, adopting LLM-based security-by-design and vulnerability scanning, accelerating cyber hygiene measures, reducing the attack surface, implementing security in depth, updating and testing crisis response arrangements and raising security standards among third parties.
- ✓ **Smaller bank:** Key priorities include making vulnerabilities visible to technical teams, ensuring active follow-up through a dedicated vulnerability manager, providing transparency for management bodies through KPIs, automating remediation, patching and incident response, adapting cyber response playbooks and crisis exercises and using AI-based vulnerability scanning and remediation in in-house software development.
- ✓ **Common features:** Vulnerability and patch management, automation, cyber response, exercises and the management of third parties and service providers are key priorities across institutions.

Guiding principles: Scalable patch and incident management processes, a sound understanding of IT infrastructure and assets, core cybersecurity measures, the use of AI technologies for cyber defence and enhanced and well-rehearsed resilience structures are at the heart of the approach.

Strategic approach: a shared vision for 2027+ and controlled implementation

Frontier AI risks should be integrated into regular ICT risk management on a permanent basis by 2027+. In line with ECB expectations, this includes reviewing risk appetite, tolerance thresholds, control measures, resource allocation decisions and management body oversight. The aim is to create a scalable control environment that brings together operational measures, makes them measurable and embeds them sustainably within the organisation. On 7 July 2026, the European Central Bank set out its expectations in more concrete terms in a letter to the management bodies of banks falling under the Single Supervisory Mechanism (SSM).

Align existing initiatives: Existing measures should be brought together within a shared 2027+ target state, defining clear implementation waves, responsibilities, success criteria and risk-based decision-making paths.

- ✓ **Scaling:** Pilot projects and individual measures should be transformed into repeatable standard processes based on risk considerations.
- ✓ **Enhance risk management:** Risk tolerance frameworks, tolerance thresholds and control measures should be adapted to the increased speed and scaling of AI-driven threats.
- ✓ **Steer and monitor:** Progress should be tracked through clear KPIs, escalation paths, resource allocation decisions and regular updates to management.

Key takeaway: Frontier AI risks require enhanced ICT risk management, with an adapted risk appetite, clearly defined tolerance thresholds and measurable management oversight.

PARTNER ORGANISATIONS

bankenverband bankenakademie

BVbank-verlag