ensono®

WHITE PAPER | INSURANCE LEADERSHIP

Attack at Machine Speed

AI-Enabled Threats to Insurance Mainframe Environments, and the Assessment Discipline That Answers Them

Jane Bradshaw, Ensono Security Product Management
Phil Karecki, Ensono Field CTO
Chandlor Mullins, Ensono Security Product Management

Executive Summary

Mainframes have long carried an unwritten security control: the scarcity of people who understand them. Compromising core insurance systems required fluency in z/OS, RACF, ACF2, Top Secret, JCL, CICS, IMS and Db2, knowledge concentrated in a small professional community. That scarcity was never a designed control, but it functioned as one. Artificial intelligence is now dissolving it.

This is not a speculative claim. Between late 2025 and mid-2026 the public record moved from theory to documented practice. Google's threat intelligence team identified the first malware families that call a large language model during execution to rewrite themselves and generate collection commands.[1] Anthropic published a detailed account of an espionage campaign in which its model executed an estimated 80 to 90 percent of tactical operations independently across roughly thirty targeted entities, sustaining multiple operations per second.[2] CrowdStrike measured average eCrime breakout time (initial access to lateral movement) falling to 29 minutes, with a fastest observed case of 27 seconds, alongside an 89 percent year-over-year increase in attacks attributed to AI-enabled adversaries.[3]

For insurers, the consequence is specific. The mainframe holds personal, commercial and specialty policy administration, claims adjudication and loss reserving, premium billing and collections, agent and broker commission accounting, reinsurance cession and recovery, annuity and life valuation, and decades of policyholder records. Its security state is expressed in platform-native constructs and is typically reviewed on a periodic, human-paced cycle. An adversary using AI to graph privilege relationships, correlate SMF activity, and construct cross-platform attack paths is not constrained to that cycle. The asymmetry is no longer about who knows more about the mainframe. It is about who can analyze its relationships faster.

---

[1]Google Threat Intelligence Group, "Advances in Threat Actor Usage of AI Tools", https://cloud.google.com/blog/topics/threat-intelligence/threat-actor-usage-of-ai-tools

[2]Anthropic, "Disrupting the first reported AI-orchestrated cyber espionage campaign" (GTG-1002), https://www-cdn.anthropic.com/d7dd50dd1185f59be051b307150d877f2b82bd2c.pdf

[3]CrowdStrike, "2026 Global Threat Report" key findings, https://www.crowdstrike.com/en-us/blog/crowdstrike-2026-global-threat-report-findings/;

CyberScoop, "Attack breakout time keeps shrinking, CrowdStrike annual report finds", https://cyberscoop.com/crowdstrike-annual-global-threat-report-attack-breakout-time/

What Security Leaders Should Now Assume

1. Specialist knowledge is no longer a barrier to entry. Reconnaissance, privilege analysis and script generation that once required years of platform experience can now be substantially automated.
2. Attackers think in paths, not findings. Individually low-severity items (a dormant privileged ID, a surrogate authority, a permissive dataset default) are the raw material of a high-severity chain.
3. The mainframe boundary is porous by design. Federated identity, shared service accounts and distributed-to-mainframe trust relationships mean many mainframe attack paths begin in Windows, cloud or SaaS estates.
4. Insiders inherit the same leverage. AI lowers the technical bar for an employee, contractor or third party who already holds legitimate access.
5. Periodic compliance validation is now insufficient as a control, and increasingly insufficient as evidence. Regulators are explicitly naming AI-amplified vulnerability discovery as a supervisory concern.

The last point is the one that changes budget conversations. In May 2026 the New York Department of Financial Services issued guidance on heightened cybersecurity risks associated with frontier AI models, describing the risk as tools that "amplify the potency, scale, and speed of identifying vulnerabilities and exploits in information systems" and directing regulated entities, insurers included, toward four focus areas: vulnerability management, third-party management, secure programming practices, and heightened monitoring and prompt reporting.[4] The letters create no new legal requirement, but NYDFS states plainly that they will inform its supervisory and enforcement expectations.

Read This Before The Threat Sections

The calibration that matters.

The most cyber-capable frontier models are, at the time of writing, access-restricted and used largely for defensive research. Ireland's National Cyber Security Centre stated in April 2026 that there was "no indication that a comparable autonomous vulnerability discovery capability is available to threat actors at this time." Rapid7 reached a similarly measured conclusion, framing the near-term output as viable attack paths and proof-of-concept exploits rather than confirmed breaches, and noting that "for most organizations, this is not a market-wide operational shift today." This paper does not argue that an AI system has broken into a named insurer's mainframe. It argues something harder to dismiss: the capability curve is public, measurable, and steepening, and the defensive work it requires takes longer to stand up than the capability takes to spread.

[4]Mayer Brown, "NYDFS Issues Two Industry Letters on Responding to a Heightened Cybersecurity Threat Environment, Including Frontier AI Models", https://www.mayerbrown.com/en/insights/publications/2026/05/nydfs-issues-two-industry-letters-on-responding-to-a-heightened-cybersecurity-threat-environment-including-frontier-ai-models

The Response This Paper Recommends

Insurers do not need a new category of tool. They need to apply the same analytical technique defensively that adversaries are applying offensively: graph the environment, model it from the attacker's perspective, and continuously identify the shortest exploitable path to policyholder, claims and payment data, across the mainframe and the distributed estate together, on a cadence measured in days rather than quarters, with named human accountability for every remediation decision.

Shift	From → To
Unit of analysis	Individual misconfiguration findings → complete attack paths from an entry point to a business-critical data asset
Scope boundary	Mainframe reviewed separately from distributed and cloud → one correlated identity and privilege graph across all three
Cadence	Annual or biennial audit-driven review → continuous assessment with risk-tiered remediation windows
Success measure	Findings closed → exploitable paths eliminated and independently retested
Accountability	Tool output distributed to platform teams → named overseer of the remediation process with a documented evidence trail
Output audience	Platform-native reports → enterprise risk language boards, regulators and underwriters can consume

Table 1. The six shifts that define an AI-era mainframe vulnerability program.

1. The Eroding Advantage

Historically, mainframes benefited from a significant security advantage: specialized knowledge. Attackers required deep expertise in z/OS, RACF, ACF2, Top Secret, JCL, CICS, IMS and Db2 to compromise core systems. That expertise took years to acquire, was rarely documented publicly in exploitable detail, and could not be bought at scale. It made the mainframe an unattractive target relative to the distributed estate, where commodity tooling and abundant tutorials lowered the cost of an attempt.

Generative AI, autonomous agents and AI-assisted offensive security tooling change the economics of that calculation in four ways. They accelerate research against sparse and technical documentation. They automate analysis of large relationship sets. They identify patterns across datasets no human team would read in full. And they reduce the expertise required to translate a finding into a working procedure. For property and casualty, life, annuity and specialty carriers that rely on mainframes to rate and issue policies, adjudicate and pay claims, process premium and financial transactions, and maintain customer records, this changes the threat equation considerably.

It is important to be precise about the mechanism, because imprecision here undermines credibility with platform teams. AI does not magically "hack RACF." There is no model that issues a command and receives SPECIAL authority. What AI does extremely well is analyze enormous numbers of relationships (permissions, group connections, dataset access, started tasks, surrogate authorities, job ownership, federated identity mappings) far faster than humans can, and then reason about how those relationships combine. On a platform whose security posture is fundamentally a very large relationship graph accumulated over decades, that is the capability that matters.

> The mainframe's defensive strength has always been the rigor of its controls, not the obscurity of its internals. The controls remain excellent. It is the obscurity that is expiring. Obscurity was doing more work than most estates accounted for.

Why Insurance Carries Concentrated Exposure

Four characteristics make insurance mainframe estates disproportionately attractive. First, data concentration: policy administration, claims, billing and payment systems hold regulated personal, medical and financial data in the same platform, including Social Security numbers, tax identification numbers, driver's license numbers, vehicle identifiers, property schedules, loss histories and banking information. Second, accumulated change: policy books acquired through mergers, run-off books carried forward from discontinued lines, multiple policy masters inherited from acquired carriers, and conversion remnants accumulate alongside decades of access grants to create a security configuration whose current state no single person fully holds. Third, operational criticality: the platform runs real-time transaction processing (nightly policy cycles, claims cycles, statutory close, renewal seasons and catastrophe response periods) whose interruption is itself a reportable business event, which raises the bar for any change and lengthens remediation windows.

Fourth, and easy to underweight in a platform-centric review: insurance runs on other people's hands. Independent agents, brokers, MGAs, TPAs, reinsurers, offshore processing partners and print/mail vendors routinely hold credentialed access to mainframe-resident data as a condition of doing business. Identity and third-party risk are not adjacent to the mainframe security question for an insurer; they are inside it.

Exposure also varies by line of business, and a program that treats the estate as undifferentiated risk misses that variation. Personal lines carry high-volume, moderately sensitive consumer data across a large policyholder base. Commercial lines carry lower volume but materially higher-value exposure, concentrated in

fewer, larger accounts. Specialty insurance typically carries the least standardized data model and the most bespoke system customization, which complicates both discovery and remediation. Life insurance and annuities carry long-duration, high-sensitivity data, including Social Security numbers, beneficiary designations and banking details, tied to policies that can remain open for decades, with a correspondingly longer window in which a compromise can go undetected. A single enterprise-wide risk rating obscures this; the assessment discipline this paper recommends should be capable of expressing risk at the line-of-business level.

The industry's own survey data reflects the resulting concern. BMC's 2026 mainframe survey found security and compliance leading mainframe organizational priorities for the third consecutive year, cited by 67 percent of respondents, a six-point increase year over year.[5] In BMC's financial-services cut, 58 percent of organizations reported using privileged-user monitoring, 47 percent privileged access management, 45 percent multi-factor authentication and 46 percent external penetration-testing services, ahead of other sectors, but well short of universal adoption of any single control.[6] Arcati's 2025 user survey found 53 percent of mainframe users expressing concern about mainframe security, with data breaches, unplanned downtime, ransomware, compromised credentials and insider threats as the top five cited risk areas, and 21 percent reporting insufficient cybersecurity monitoring.[7] In a separate modernization survey, 69 percent of IT leaders named data security as their principal modernization concern.[8]

Read together, these figures describe an estate where leaders understand the risk and have deployed real controls, but where monitoring coverage, privileged-access discipline and independent testing are uneven. That unevenness is precisely what path-based analysis exploits.

---

[5]BMC, "BMC Survey Confirms Mainframe Platform Innovation" (2026 Mainframe Survey), https://www.bmc.com/newsroom/releases/bmc-survey-confirms-mainframe-platform-innovation.html

[6]BMC, "Mainframe Survey Financial Services: Key Takeaways", https://www.bmc.com/blogs/mainframe-survey-financial-services-key-takeaways/

[7]Arcati, "Mainframe Navigator 2025 User Survey", https://planetmainframe.com/wp-content/uploads/arcati/Arcati-Mainframe-Navigator-2025-User-Survey.pdf

[8]Planet Mainframe / Rocket Software modernization survey findings, https://planetmainframe.com/2026/03/new-survey-findings-new-flashsystem-portfolio-and-more/

2. What Changed, and When

The argument in this paper rests on a public evidence base assembled over roughly twelve months. The timeline below is deliberately conservative: it includes only items published by the organization that observed them, by a government body, or by established trade press reporting on a named primary document.

Date	Event	Why it matters to an insurer
Oct 2025	OpenAI reports disrupting more than 40 policy-violating networks since Feb 2024	Establishes sustained, industrialized threat-actor use of commercial models. OpenAI's own framing is that actors "bolt AI onto old playbooks to move faster" rather than gain novel capability.
Nov 2025	Google GTIG identifies PROMPTFLUX and PROMPTSTEAL, the first malware observed calling an LLM during execution	Malware that rewrites itself hourly to evade detection invalidates signature-led assumptions. PROMPTSTEAL, used by a Russian state-backed actor, generated its collection commands at runtime.
Sep-Nov 2025	Anthropic documents the GTG-1002 campaign: model executed an estimated 80-90% of tactical operations autonomously against ~30 entities	First published account of an intrusion campaign largely executed without human intervention at scale, at sustained rates of multiple operations per second.
Feb 2026	CrowdStrike reports 29-minute average eCrime breakout time (27 seconds fastest), 89% rise in AI-enabled adversary attacks, 42% rise in pre-disclosure zero-day exploitation	Defines the clock. Any control whose response time is measured in weeks is outside the engagement window.
Apr 2026	Access-restricted frontier model with substantial autonomous vulnerability-discovery capability made available to a limited partner set; UK AISI publishes independent capability evaluation	Demonstrates that autonomous discovery of viable attack paths is now technically achievable, while remaining access-controlled.
Apr 2026	Ireland's NCSC states there is "no indication that a comparable autonomous vulnerability discovery capability is available to threat actors at this time"	The calibrating counterweight. The capability exists; broad adversarial availability is not established.

Date	Event	Why it matters to an insurer
May 2026	CISA, NSA and Five Eyes partners publish "Careful Adoption of Agentic AI Services", first joint guidance scoped to agentic AI	Names privilege creep and expanded attack surface as primary agentic risks and recommends per-agent identity, short-lived credentials and a human-approval matrix.
May 2026	NYDFS issues industry letters on a heightened threat environment and on frontier AI model risk	Directly applicable supervisory expectation for insurers and other regulated entities in New York.
Jun 2026	EO 14409 signed; CISA issues BOD 26-04 replacing patch-everything with four-variable risk prioritization	Establishes the emerging federal standard for what "prioritized remediation" means, including a three-day window at the top tier.
Jul 2026	Frontier labs disclose that their own models autonomously accessed third-party environments during testing, using basic techniques such as weak passwords	Underlines that the first casualties of capable models are not exotic vulnerabilities but ordinary hygiene failures.

Table 2. Twelve months of public evidence. Sources are cited in full in the sections that follow and in the reference list.

What The Evidence Supports And Does Not Support

Precision here protects the business case. The following claims are supported by the public record: threat actors are using commercial AI models across the full attack lifecycle, including reconnaissance, lure creation, command-and-control development and exfiltration;[9] malware calling an LLM at runtime has been observed in live operations;[10] at least one campaign has been executed largely autonomously by a model under high-level human direction;[11] and measured intrusion velocity is increasing sharply.[12]

The following claims are *not* supported and should not be used in an internal business case: that a frontier model of the most capable class has been used in a confirmed attack against a named organization; that mainframe-specific autonomous exploitation has been observed in the wild; or that a specific numeric probability can be attached to either. Rapid7's framing is the appropriate register: the near-term output of these capabilities is viable attack paths and proof-of-concept exploits, and for most organizations this is not yet

[9]Google Threat Intelligence Group, "Advances in Threat Actor Usage of AI Tools", https://cloud.google.com/blog/topics/threat-intelligence/threat-actor-usage-of-ai-tools

[10]Google Threat Intelligence Group, "Advances in Threat Actor Usage of AI Tools", https://cloud.google.com/blog/topics/threat-intelligence/threat-actor-usage-of-ai-tools

[11]Anthropic, "Disrupting the first reported AI-orchestrated cyber espionage campaign" (GTG-1002), https://www-cdn.anthropic.com/d7dd50dd1185f59be051b307150d877f2b82bd2c.pdf

[12]CrowdStrike, "2026 Global Threat Report" key findings, https://www.crowdstrike.com/en-us/blog/crowdstrike-2026-global-threat-report-findings/

a market-wide operational shift.[13] Overstating the case invites a credibility challenge that the underlying argument does not need.

There is also a defensive mirror worth noting in board discussion. Security vendors are formalizing the same techniques for legitimate use. Rapid7, for example, has published its own multi-agent AI red-teaming architecture with tiered safety controls including scope enforcement, action classification and human-in-the-loop by default.[14] The technique is not inherently adversarial. The question for an insurer is which side applies it to their estate first.

[13]Rapid7, "Project Glasswing: What Security Leaders Should Know and Do Now", https://www.rapid7.com/blog/post/ai-what-project-glasswing-means-for-security-leaders/

[14]Rapid7, "Formalizing Red Teaming Offensive Methodology as a Multi-Agent AI Architecture", https://www.rapid7.com/blog/post/so-red-teaming-offensive-methodology-multi-agent-ai-architecture/

3. Eight AI-Enabled Attack Vectors Against z/OS Environments

The eight vectors below describe how AI changes attacker capability against a mainframe estate. Each is presented with the mechanism, its specific relevance to insurance workloads, and the defensive response it demands. None depend on a novel exploit or an undisclosed vulnerability. All depend on analyzing relationships and history that already exist in the environment.

Vector	What AI changes	Defensive response
Privilege escalation analysis	Graphs entitlement relationships across RACF, ACF2 or Top Secret to find the shortest path from a low-privilege ID to enterprise authority	Continuously assess entitlements, privileged access paths, dormant IDs and separation-of-duty conflicts using graph-based analytics
Attack path mapping	Chains individually low-severity findings into a complete route to a business-critical data asset	Assess how combinations of low-risk findings aggregate into high-impact business risk; report at path level
Accelerated reconnaissance	Analyzes SMF records, job schedules, batch dependencies and system logs at machine speed to locate high-value targets	Monitor anomalous access patterns continuously; maintain a current inventory of sensitive data repositories
AI-assisted insider threat	Removes the skill barrier for a user who already holds legitimate access	Behavioral monitoring, privileged user analytics, enforced least privilege, and time-bound elevation
Targeted social engineering	Produces highly tailored pretexts against named mainframe and security personnel by combining public and previously exfiltrated data	Phishing-resistant MFA for privileged roles, privileged activity monitoring, role-specific awareness training, verified out-of-band change authorization
Configuration drift detection	Reviews years of access changes, exceptions and audit findings to find weaknesses created incrementally	Baseline RACF, ACF2 and Top-Secret configuration and continuously assess drift from the approved baseline
Data discovery and exfiltration planning	Locates regulated data and ranks policy, claims, disbursement, commission and reinsurance datasets by financial, regulatory and operational impact	Classify sensitive data, encrypt critical datasets, monitor access patterns, continuously validate access controls

Vector	What AI changes	Defensive response
Malware and automation generation	Lowers the barrier to producing automation scripts, modified tooling and exploitation workflows, compressing adaptation cycles	Behavioral and anomaly detection plus continuous vulnerability assessment, rather than reliance on signature-based controls

Table 3. The eight vectors at a glance. Each is expanded below.

3.1 AI-Powered Privilege Escalation Analysis

An attacker can use AI to analyze RACF, ACF2 or Top-Secret security structures and identify privilege relationships that would be difficult for a human analyst to uncover manually. The classes of finding are familiar to any mainframe security team: excessive administrative privileges, dormant privileged accounts, segregation-of-duty violations, inherited permissions, shared service accounts and misconfigured surrogate authorities. Published security research has documented the specific escalation mechanics involved, including abuse of SURROGAT permissions, group-level SPECIAL, OPERATIONS and AUDITOR attribute relationships, IRR.PASSWORD.RESET facility-class abuse, and z/OS UNIX superuser escalation through BPX.SUPERUSER or UID(0) assignment.[15]

What changes is throughput. Rather than reviewing thousands of access relationships individually, an attacker can identify the shortest path from a low-level user account to highly privileged access in minutes. The finding was always discoverable. It was the cost of discovery that protected it.

Defensive response.
Continuously assess entitlements, privileged access paths, dormant IDs and separation-of-duty conflicts using the same graph-based analytics an attacker would use. Assessment scope must include group nesting and indirect authority, not only directly granted permissions.

3.2 AI-Driven Attack Path Mapping

Modern attackers increasingly use graph analytics and AI models to identify complete attack chains rather than isolated vulnerabilities. A seemingly low-risk configuration issue becomes critical when combined with an unused privileged account, excessive dataset permissions, weak service account controls, inadequate MFA enforcement, or a trust relationship between the distributed and mainframe environments. AI excels at surfacing these hidden relationships. Instead of finding one vulnerability, the attacker identifies the fastest route to sensitive claims, customer, payment or policy data.

This is the single most consequential shift in the paper, because it invalidates severity-based triage as a complete strategy. A program that defers every medium and low finding is, in path terms, deferring the majority of the material risk. The Five Eyes agentic AI guidance makes a structurally identical point about autonomous systems, identifying privilege accumulation and expanded attack surface as leading risk

[15]Kaspersky Securelist, "Deconstructing RACF in z/OS and uncovering security issues", https://securelist.com/zos-mainframe-pentesting-resource-access-control-facility/116873/

categories and recommending that high-impact action approval be defined by system designers rather than delegated to the agent.[16]

Defensive response.
Conduct attack-path analysis that evaluates how multiple low-risk findings combine into high-impact business risk. Prioritize remediation by the number of paths a fix eliminates, not by the severity score of the individual finding.

3.3 AI-Accelerated Mainframe Reconnaissance

Reconnaissance against a mainframe has traditionally been labor intensive and highly specialized. AI now enables rapid analysis of SMF records, user activity patterns, network flows, job schedules, batch processing dependencies and system logs. Patterns that previously required weeks of manual review can be surfaced automatically, producing faster identification of high-value datasets, privileged users, critical business applications, backup locations and data replication targets.

Backup and replication targets deserve particular attention in an insurance context. They frequently hold the same regulated data as production under weaker access control and lighter monitoring, and they are often excluded from the scope of platform security reviews on the grounds that they are not production systems.

The insurance-specific target list extends further than the platform-security literature typically accounts for: bordereaux files, reinsurance cession files, bureau submissions, actuarial data feeds, catastrophe modeling extracts and modernization-project data extracts all move regulated data outside the production security perimeter as a matter of routine business process, often on a recurring schedule an attacker can learn.

Defensive response.
Monitor anomalous access patterns continuously and maintain current visibility into every repository holding sensitive data, including backups, replicas, test copies and data extracts feeding analytics platforms.

3.4 AI-Assisted Insider Threats

AI lowers the skill requirements for a malicious insider. A user with legitimate access but limited technical expertise can use AI tools to generate scripts, analyze permissions, identify sensitive datasets, discover access pathways and develop automated extraction techniques. The population of concern is therefore not only external attackers but also employees, contractors and third parties who now hold capabilities previously reserved for experienced security professionals, including claims adjusters, claims examiners, underwriting support staff, offshore BPO partners, TPAs and systems integrators, each of whom may already hold standing access to policy or claims data as a condition of their role.

The IBM Cost of a Data Breach 2025 findings are relevant here in an unexpected way. Twenty percent of breached organizations reported shadow-AI-related security incidents, and organizations with high shadow-AI exposure averaged 4.74 million dollars in breach cost against 4.07 million for those with low or no exposure.

---

[16]CISA, NSA and Five Eyes partners, "Careful Adoption of Agentic AI Services", https://www.cisa.gov/resources-tools/resources/careful-adoption-agentic-ai-services;

Crowell & Moring, "American and Allied Cyber Agencies Issue First Joint Guidance on Securing Agentic AI", https://www.crowell.com/en/insights/client-alerts/american-and-allied-cyber-agencies-issue-first-joint-guidance-on-securing-agentic-ai

Ninety-seven percent of organizations that suffered an AI-related security incident lacked proper access controls on their AI systems.[17] The insider risk and the AI governance risk are the same risk viewed from two directions.

Defensive response.
Implement continuous behavioral monitoring, privileged user analytics and enforced least privilege. Extend acceptable-use governance and access control to the AI tooling your own staff and contractors use, and inventory it.

3.5 AI-Enhanced Social Engineering Against Mainframe Personnel

Mainframe administrators remain among the most privileged individuals in an insurance organization, and they are a small, identifiable population. AI enables attackers to generate highly customized phishing, business email compromise and impersonation attacks against RACF administrators, security teams, infrastructure engineers, operations staff and third-party support personnel, combining publicly available information with internal organizational data obtained through prior compromise.

Compromised credentials remain a material initial vector. Broadcom cites credential-based attacks costing organizations 4.67 million dollars annually and serving as the initial attack vector 10 percent of the time and has removed the legacy no-password (NOPW) option in ACF2 and Top-Secret version 17.0 on the grounds that it had been exploited for user ACIDs.[18] The relevant point for a security program is that legacy authentication exceptions carried forward for operational convenience are the exact material an AI-assisted campaign is looking for.

Insurance also supplies its own calendar of high-pressure pretexts, and an AI-assisted campaign can time itself to it: annual statement filing, reserve certification, insurance department examinations, catastrophe response periods and conversion weekends all create windows in which a request framed as urgent, senior and time-boxed is more likely to be honored without the verification it would otherwise receive.

Defensive response.
Enforce phishing-resistant MFA on every privileged mainframe role without exception, monitor privileged account activity, deliver role-specific awareness training to the platform population, and require verified out-of-band authorization for privilege changes and password resets.

3.6 AI-Powered Identification Of Configuration Drift

Insurance organizations often manage decades of accumulated change across mainframe environments. AI can review years of access control changes, dataset permissions, security exceptions, audit findings and system configurations and identify weaknesses created through gradual drift. What appears in isolation to be a legitimate business exception can constitute significant exposure when viewed against the full history of

[17]IBM, "Cost of a Data Breach Report 2025", https://www.ibm.com/reports/data-breach;

IBM Think, "2025 Cost of a Data Breach: navigating AI", https://www.ibm.com/think/x-force/2025-cost-of-a-data-breach-navigating-ai

[18]Broadcom, "Modernize Security with ACF2 and Top-Secret Version 17.0", https://news.broadcom.com/mainframe-software/modernize-security-with-acf2-and-top-secret-version-17-0

changes around it. In insurance environments, the recurring sources of that drift are predictable: acquired books of business, conversion projects, product launches, state filings and system mergers each leave behind access grants and configuration exceptions that were justified for a defined period and never revisited once that period ended.

The recurring misconfiguration classes are well documented in the mainframe security community and have remained remarkably stable: excessive access to APF-authorized libraries, privileged users not enrolled in MFA, weak password encryption, started-task and batch IDs with excessive access, sensitive datasets with a default access greater than NONE, excessive z/OS UNIX superuser privilege, and IDs with unchanged non-expiring passwords.[19] The persistence of this taxonomy is itself the finding. These are not unknown weaknesses; they are known weaknesses that periodic review has not eliminated.

Defensive response.
Establish an approved security baseline for RACF, ACF2 and Top Secret, and continuously assess drift from it. Require every exception to carry an owner, a business justification, a compensating control and an expiry date.

3.7 AI-enabled data discovery and exfiltration planning

Large language models are effective at finding patterns in unstructured and semi-structured data. An attacker can use AI to identify regulated data, locate customer information, discover sensitive financial records, map data relationships and rank high-value targets, determining quickly which datasets would generate maximum financial, regulatory or operational impact. In an insurance estate that means claims master files, adjuster notes, the policy master, commission files, agency banking information, treaty terms and catastrophe exposure data, each of which carries a different regulatory and financial profile if exposed. For insurers, exposed policyholder, claims, payment and personally identifiable information remains an attractive target, and the regulatory consequence of its loss is defined in statute rather than negotiated.

Under the NAIC Insurance Data Security Model Law, a licensee must notify the Commissioner no later than 72 hours after determining that a cybersecurity event has occurred, submit an annual written compliance certification by 15 February, and retain supporting records for five years.[20] An attacker who has already ranked your datasets by regulatory impact understands your notification exposure better than an internal team working from an out-of-date data inventory.

Multi-state carriers carry the additional problem of jurisdictional variation: notification thresholds, timelines and required content differ by state, so the same exposed dataset can trigger materially different obligations depending on where the affected policyholders reside, a distinction an out-of-date data inventory rarely captures and an attacker has no reason to respect.

Defensive response.
Classify sensitive data at the dataset level, encrypt critical datasets, monitor access patterns against classification, and continuously validate that access controls match the classification. Maintain the data inventory as an operational artifact, not an audit deliverable.

[19]Guide Share Europe UK, mainframe External Security Manager vulnerability taxonomy (2020), https://conferences.gse.org.uk/2020/presentations/1AZ.pdf

[20]NAIC, Insurance Data Security Model Law (MDL-668), https://content.naic.org/sites/default/files/model-law-668.pdf

3.8 AI-Generated Malware & Automation

Sophisticated mainframe-native malware remains rare, and this paper does not claim otherwise. The material change is that AI lowers development barriers for automation scripts, modifications to existing tools, attack playbooks, custom exploitation workflows and vulnerability research. The concern is less about fully autonomous mainframe malware and more about the speed at which attackers can adapt and evolve technique.

The distributed-estate precedent is instructive. Google's threat intelligence team documented PROMPTFLUX, which calls a commercial model through a "Thinking Robot" module to obtain new antivirus-evasion code, with one variant instructing the model to rewrite the malware's entire source on an hourly basis to evade detection, and PROMPTSTEAL, which queries an open-source model to generate its data-collection and exfiltration commands in live operations.[21] Neither targets z/OS. Both demonstrate that the adaptation cycle can now be shorter than the detection-signature cycle.

> **Defensive response.**
> Weight behavioral detection, anomaly monitoring and continuous vulnerability assessment above signature-based controls. Assume that any control whose efficacy depends on a known indicator has a shorter useful life than it did two years ago.

[21]Google Threat Intelligence Group, "Advances in Threat Actor Usage of AI Tools", https://cloud.google.com/blog/topics/threat-intelligence/threat-actor-usage-of-ai-tools

4. How AI Finds Privilege Escalation: Five Worked Examples

Privilege escalation occurs when a user with limited access gains higher-level privilege than intended. In a mainframe environment the challenge is rarely finding a single obvious misconfiguration. It is identifying how multiple individually harmless permissions combine to create a route to elevated access. Security teams have traditionally analyzed RACF groups, user IDs, dataset permissions, surrogate permissions, started-task definitions, program access controls, shared accounts and trust relationships, often reviewing thousands or millions of relationships manually. The five examples below illustrate what changes when that analysis is performed by a system that can hold the whole graph at once.

Example 1: Hidden Group Inheritance

Consider a claims processor with an ordinary user ID. On the surface the account holds no privileged access. Graph analysis reveals otherwise:

Claims user ID

↓

Member of Group A

↓

Group A connected to Group B

↓

Group B holds UPDATE access to sensitive datasets

↓

Dataset contains payroll job control

↓

Job executes under a privileged account

A human reviewer examining any single link would find nothing objectionable, and each grant may have been individually approved. An AI model graphing every user, group, permission, dataset and job relationship simultaneously identifies that a nominally low-privilege account can ultimately influence privileged processing. The control failure is not in any one grant. It is in the absence of any process that evaluates the composition of grants.

Example 2: Attack Path Discovery Across Low-Severity Findings

Assume an insurer's estate contains a contractor account, a dormant administrator account, an excessive surrogate authority and a weak password policy on a service ID. Individually, none would typically be rated critical. In combination:

Contractor account

↓

Access to a shared application

↓

Able to submit a batch job

↓

Batch job executes through a service ID

↓

Service ID holds surrogate authority

↓

Surrogate authority reaches security administration functions

Four low-risk findings constitute one high-risk chain terminating in security administration. This is how sophisticated threat actors increasingly reason: not in terms of individual vulnerabilities but in terms of complete paths to sensitive assets. It is also the strongest available argument against a remediation policy that closes only high and critical findings.

Example 3: Peer-Relative Excessive Privilege

Many insurance organizations accumulate permissions over decades. Comparing a user against the statistical behavior of their peer group surfaces outliers that absolute rules miss:

Peer group baseline (500 claims analysts)	One outlier account
READ access to claims datasets	SPECIAL authority
No security administration authority	OPERATIONS authority
No dataset ALTER access	Dataset ALTER access

Table 4. Peer-relative analysis identifies the outlier immediately; rule-based review frequently does not.

An analyst reviewing this account in isolation may see nothing wrong, because the grant was made years ago, was legitimate at the time, and was never removed when the role changed. Peer-relative analysis makes the anomaly obvious without requiring anyone to remember the original justification.

Example 4: Dormant Privileged Accounts

Attackers favor dormant accounts because they escape scrutiny; no user notices anomalous behavior on an account no user is watching. Correlating login history, SMF activity, entitlement data, job execution records and ticketing data produces a contextual finding rather than a bare inventory entry:

Attribute	Observed value
User ID	ADMINSUPP01
Last interactive login	18 months ago
Held privileges	SPECIAL, OPERATIONS, AUDITOR
Account state	Enabled; no expiry set
Associated change record	None in current ticketing system
Contextual conclusion	Account holds enterprise-level privilege, has not been used in 18 months, remains enabled, and constitutes a high-value privilege escalation target

Table 5. The value is in the synthesized conclusion, not the individual attributes, each of which was already available.

That level of contextual analysis significantly reduces investigation effort, and more importantly it converts a low-priority hygiene item into an explicit, evidenced risk statement that a remediation process can act on.

Example 5: Mapping Lateral Movement Across The Platform Boundary

One of the most consequential capabilities is identifying how an attacker can move through an environment after initial compromise. Cross-domain correlation may reveal:

Windows account compromised via phishing

↓

Active Directory group membership

↓

Federated identity mapping

↓

Mainframe user ID

↓

Privileged RACF group

↓

Claims database

Historically, security teams reviewed distributed and mainframe access separately and organized their teams accordingly. AI correlates permissions across both and identifies paths that cross the platform boundary. This matters because many breaches begin in cloud or Windows environments before expanding into critical transaction systems. It also matters organizationally: a path that crosses a team boundary has no single owner, and unowned paths do not get closed.

The structural implication.
If the attack path crosses the mainframe boundary, the assessment scope must cross it too. An assessment that stops at the platform edge cannot see the paths that matter most, and neither can two separate assessments whose findings are never correlated.

5. From Permission Review to Attack-Path Analysis

The difference between a traditional review and a path-based assessment is best expressed as a difference in the question being asked. A traditional security review answers: "What permissions does this user have?" Path-based analysis answers: "What is the fastest route from this user to our most sensitive data?" The first question produces an inventory. The second produces a decision.

> AI does not simply identify misconfigured permissions. It identifies how an attacker could chain privileges, trust relationships, dormant accounts, service identities and configuration weaknesses to reach critical policyholder, claims and financial data. Used defensively, it lets an insurer discover those privilege-escalation paths before an adversary applies the same technique from the outside.

Dimension	Traditional permission review	Attack-path analysis
Question asked	What access does this identity hold?	What is the shortest exploitable route to a critical data asset?
Unit of output	Findings, listed by severity	Paths, ranked by business impact and exploitability
Treatment of low findings	Deferred or accepted individually	Evaluated for their contribution to a chain
Scope	Usually one platform, one team	Correlated across mainframe, distributed, cloud and identity provider
Analytical basis	Rules and policy comparison	Relationship graph across identities, groups, datasets, jobs, privilege and trust
Cadence achievable	Periodic, effort-bound	Continuous, compute-bound
Remediation guidance	Close the finding	Cut the path: often one change removes many findings from relevance
Board-level narrative	"We closed 412 findings this quarter"	"We eliminated the three shortest routes to the claims master file and independently retested them"

Table 6. The analytical shift, and why it changes both remediation prioritization and executive reporting.

The practical benefit is efficiency as much as coverage. Because paths converge, a small number of changes often eliminates a disproportionate share of exploitable routes: cutting a single over-permissioned surrogate authority may invalidate dozens of chains. That is a materially stronger argument to bring to a platform team protective of its change windows than a list of individually unremarkable findings.

6. The Control Response: Eight Commitments

The eight commitments below convert the vector analysis into a program. Each is expressed as a control, an owner, a cadence and, critically, an evidence artifact, because a control that produces no artifact cannot be reported to a regulator, a board or an underwriter.

Commitment	Control	Cadence	Evidence artifact
1. Baseline the graph	Full entitlement and relationship inventory across ESM, z/OS UNIX, started tasks, batch IDs and federated identity mappings	Establish once; refresh continuously	Versioned baseline with named owner and date
2. Assess paths, not findings	Attack-path analysis from defined entry points to defined critical data assets	Continuous, with formal review monthly	Ranked path register with business-impact statement per path
3. Eliminate standing privilege	Remove or time-bound SPECIAL, OPERATIONS, AUDITOR and surrogate authority; require justification and expiry for every exception	Quarterly recertification	Exception register with owner, compensating control and expiry date
4. Close the dormancy window	Automated detection and disablement of unused privileged IDs, including service and started-task IDs	Continuous detection; 30-day disablement threshold	Dormant-ID report with disposition per account
5. Enforce phishing-resistant MFA	No privileged mainframe role exempt; retire legacy no-password and non-expiring password options	Continuous enforcement; exceptions reviewed monthly	MFA coverage report by privileged role, with exception list
6. Assess drift continuously	Automated comparison of live ESM configuration against the approved baseline	Continuous	Drift report with change attribution and remediation status

Commitment	Control	Cadence	Evidence artifact
7. Test independently	Penetration testing and path validation performed by a party organizationally separate from the team that manages the target systems	At least semi-annual, plus event-triggered	Independent test report and retest confirmation
8. Oversee remediation, do not absorb it	A named overseer accountable for driving the remediation process to closure, with remediation executed by the accountable platform tower	Weekly operational review; monthly governance review	Remediation register with owner, tier, due date, closure basis and independent verification

Table 7. Eight commitments, each with a defined evidence artifact.

Why Remediation Windows Should Be Risk-Tiered

Uniform remediation SLAs fail under AI-era conditions: they are simultaneously too slow for the small set of vulnerabilities an automated attacker will reach first and too aggressive for the large set that pose little practical risk. CISA's Binding Operational Directive 26-04, issued 10 June 2026, formalizes an alternative that is worth adopting as a private-sector reference model even though it binds only federal civilian agencies. It scores a vulnerability on four variables: whether the asset is publicly exposed to unauthenticated entities; whether the CVE appears in CISA's Known Exploited Vulnerabilities catalog; whether exploitation can be fully automated; and whether exploitation grants partial or total control of the asset. Where all four are true, remediation is required within three calendar days together with mandatory forensic triage for prior compromise; lesser combinations fall into 14-day or 60-day windows or defer to the next scheduled upgrade.[22]

The third variable, full automatability, is the one that connects the directive to this paper's thesis. It is an explicit regulatory acknowledgement that the speed at which an adversary can act is a first-class input to prioritization. Executive Order 14409, signed 2 June 2026, sits behind the directive, directing federal cyber-defense prioritization through binding directives, establishing an AI cybersecurity clearinghouse under Treasury to coordinate vulnerability scanning, validation and patch distribution with industry, and creating a classified benchmarking process for designating covered frontier models by cyber capability.[23]

---

[22]SecurityWeek, "CISA Directs Federal Agencies to Prioritize Security Patches Based on Risk" (BOD 26-04), https://www.securityweek.com/cisa-directs-federal-agencies-to-prioritize-security-patches-based-on-risk/;

CyberScoop, "CISA directive orders agencies to prioritize vulnerability remediation" (BOD 26-04), https://cyberscoop.com/cisa-vulnerability-remediation-directive-bod-26-04/;

FedRAMP Public Notice 0014, "FedRAMP Response to CISA BOD 26-04", https://www.fedramp.gov/notices/0014/

[23]The White House, Executive Order 14409, "Promoting Advanced Artificial Intelligence Innovation and Security", https://www.whitehouse.gov/presidential-actions/2026/06/promoting-advanced-artificial-intelligence-innovation-and-security/;

Federal Register, EO 14409 publication record (91 FR 34565), https://www.govinfo.gov/app/details/FR-2026-06-05/2026-11415;

Congressional Research Service, summary of EO 14409, https://www.congress.gov/crs-product/IF13268

Tier	Trigger	Window	Additional obligation
Tier 1	All four variables true	3 calendar days	Forensic triage for evidence of prior compromise
Tier 2	Two or three variables true	14 calendar days	Documented interim compensating control
Tier 3	One variable true, or elevated exploitability intelligence	60 calendar days	Tracked in the path register until closed
Tier 4	None of the four	Next scheduled maintenance window	Recorded with justification for deferral

Table 8. A risk-tiered remediation model adapted from the BOD 26-04 variable set for private-sector use.

7. Regulatory and Board Expectations

The regulatory environment is not slowing AI adoption in insurance security. It is accelerating it, while raising the standard of proof. Four developments define the current expectation.

7.1 NYDFS Frontier AI Guidance

On 21 May 2026 the New York Department of Financial Services issued two industry letters: guidance on measures regulated entities should consider in a heightened cybersecurity threat environment, and guidance on heightened cybersecurity risks associated with frontier AI models. The Department characterizes the risk as AI tools that "amplify the potency, scale, and speed of identifying vulnerabilities and exploits in information systems," notes that such models are "not yet broadly available" but "may become more available soon," and directs attention to four areas: vulnerability management, third-party management, secure programming practices, and heightened monitoring and prompt reporting. Both letters state that they create no new legal requirements but will inform the Department's supervisory and enforcement expectations.[24] This follows the second amendment to Part 500, whose expanded multi-factor authentication and asset-inventory provisions took effect 1 November 2025.[25]

7.2 NAIC AI Governance And Data Security

The NAIC Model Bulletin on the Use of Artificial Intelligence Systems by Insurers, adopted in December 2023, established the governance template now reproduced across a substantial share of states; a widely used legal tracker recorded roughly half of US states having adopted it in full or near-full form, with individual adoption dates spanning February 2024 to March 2025.[26] Separately, the Insurance Data Security Model Law sets the 72-hour notification obligation, the 15 February annual certification and the five-year record retention requirement noted earlier.[27] The combination is significant: an insurer using AI defensively is subject to governance expectations for that use, and an insurer suffering an AI-enabled breach is subject to a notification clock measured in hours.

State insurance departments review information-technology general controls as part of the financial examination process, which gives this exposure a supervisory channel distinct from data-breach notification. A mainframe estate with unmanaged privilege paths and unaddressed configuration drift is not only a breach risk; it is a finding waiting to surface in the next examination cycle.

---

[24]Mayer Brown, "NYDFS Issues Two Industry Letters on Responding to a Heightened Cybersecurity Threat Environment, Including Frontier AI Models", https://www.mayerbrown.com/en/insights/publications/2026/05/nydfs-issues-two-industry-letters-on-responding-to-a-heightened-cybersecurity-threat-environment-including-frontier-ai-models

[25]Greenberg Traurig, "NYDFS Final Cybersecurity Rules: MFA, Asset Inventory, and Third-Party Risk", https://www.gtlaw.com/en/insights/2025/11/nydfs-final-cybersecurity-rules-mfa-asset-inventory-and-third-party-risk

[26]NAIC, "Model Bulletin on the Use of Artificial Intelligence Systems by Insurers", https://content.naic.org/sites/default/files/cmte-h-big-data-artificial-intelligence-wg-map-ai-model-bulletin.pdf;

NAIC, artificial intelligence topic page, https://content.naic.org/insurance-topics/artificial-intelligence;

Quarles & Brady, state adoption tracker for the NAIC AI Model Bulletin, https://www.quarles.com/newsroom/publications/nearly-half-of-states-have-now-adopted-naic-model-bulletin-on-insurers-use-of-ai

[27]NAIC, Insurance Data Security Model Law (MDL-668), https://content.naic.org/sites/default/files/model-law-668.pdf

7.3 DORA And EU Operational Resilience

The Digital Operational Resilience Act entered into application on 17 January 2025 and applies to twenty types of financial entity and to ICT third-party service providers, requiring that insurers and other in-scope firms can withstand, respond to and recover from ICT disruption.[28] For groups with EU operations, mainframe vulnerability management is not a discretionary internal practice; it is part of a documented resilience obligation subject to supervisory review.

7.4 Agentic AI Guidance From CISA And Five Eyes Partners

In May 2026 CISA and the NSA, together with counterpart agencies in Australia, Canada, New Zealand and the United Kingdom, published "Careful Adoption of Agentic AI Services", the first joint international guidance scoped specifically to agentic AI. It defines a five-category risk taxonomy covering privilege, design and configuration, behavioral, structural and accountability risk, and recommends per-agent cryptographic identity with short-lived credentials, a human-approval matrix for high-impact actions defined by system designers rather than the agent, mapping of agentic risk to existing identity, segmentation and observability controls, instrumentation of agent reasoning and tool calls for audit, threat modelling before tool and data integration, and incremental fail-safe-by-default adoption beginning with low-risk use cases.[29]

The board framing.
Insurers also occupy an unusual position because they underwrite cyber risk for others while carrying it themselves; a control gap in their own mainframe estate is a claims exposure for the book as much as it is a security incident for the business, and boards should expect that framing to surface in underwriting and reinsurance conversations, not only regulatory ones. The platform's operational track record, however strong, is no longer accepted as evidence in its own right. Regulators and boards now expect the same documented proof from z/OS that they expect from cloud and distributed environments, produced on the same cadence and in the same language. The IBM Cost of a Data Breach 2025 places the average breach in the financial sector, which includes insurance, at 5.56 million dollars, against a 4.44 million dollar global average and a 10.22 million dollar US average. That exposure makes the return-on-investment case straightforward. The harder question is whether the governance architecture around the tooling is sound enough to survive scrutiny.

[28]EIOPA, "Digital Operational Resilience Act (DORA)", https://www.eiopa.europa.eu/digital-operational-resilience-act-dora_en

[29]CISA, NSA and Five Eyes partners, "Careful Adoption of Agentic AI Services", https://www.cisa.gov/resources-tools/resources/careful-adoption-agentic-ai-services;

Crowell & Moring, "American and Allied Cyber Agencies Issue First Joint Guidance on Securing Agentic AI", https://www.crowell.com/en/insights/client-alerts/american-and-allied-cyber-agencies-issue-first-joint-guidance-on-securing-agentic-ai

8. What Insurance Security Leaders Are Actually Buying

The value of an AI-driven mainframe security assessment is not automation. Automation is the mechanism, not the outcome. The outcome is the capacity to defend against adversaries who are already using AI to discover attack paths faster, analyze privilege relationships more effectively, target sensitive data more accurately, accelerate reconnaissance, scale operations with fewer resources, and reduce the expertise required to attack specialized platforms.

What it looks like on an invoice	What is actually being purchased
Automated scanning of the ESM configuration	A current, evidenced answer to "what is the shortest route from a compromised laptop to the claims master file?"
Vulnerability findings report	A ranked path register that survives challenge from platform engineering, internal audit and an external examiner
Attack-path analysis	Parity with the analytical technique adversaries are applying to the same estate
Continuous assessment subscription	A cadence that fits inside the measured intrusion window rather than the audit calendar
Independent penetration testing	Validation that a closed finding is actually an eliminated path, performed by a party with no stake in the answer
Managed remediation oversight	Named accountability, a documented decision trail, and reportable closure: the artifacts a regulator asks for

Table 9. Reframing the purchase for an executive audience.

The most effective insurers are using AI defensively to establish parity with AI-enabled attackers. The objective is no longer periodic compliance validation. The objective is continuous identification of exploitable attack paths before an adversary uses the same technique to find them first, and the production of evidence that this is happening, in language a board, a regulator and an underwriter can each act on.

9. A Ninety-Day Path

The program described in this paper does not require a multi-year transformation to begin producing value. The sequence below reflects how Ensono structures the first ninety days of a mainframe AI risk and vulnerability engagement, and it is deliberately front-loaded toward establishing evidence rather than closing findings.

Phase	Focus	Outcome at exit
Days 1-30	Scope, access and baseline. Establish the entitlement and relationship inventory; agree the critical data assets (policy master, claims master, disbursement, commission and reinsurance files) that define path endpoints; confirm onboarding readiness across the platform, identity and network towers.	A versioned baseline, an agreed critical-asset list, and a named owner per tower.
Days 31-60	Path analysis and triage. Run attack-path analysis from defined entry points; correlate mainframe with distributed and identity-provider data; tier findings against the four-variable risk model.	A ranked path register with business-impact statements and tiered remediation windows.
Days 61-90	Remediation oversight and independent validation. Drive tier 1 and tier 2 paths to closure through the accountable towers; schedule the first independent test cycle; stand up the governance reporting pack.	Tier 1 paths closed and verified; first independent test scheduled; a repeatable monthly evidence pack.

Table 10. First ninety days. Deliberately weighted toward evidence and prioritization before remediation volume.

Two design choices in that sequence are worth making explicit, because they are the ones most often compromised under delivery pressure. First, remediation oversight and remediation execution are separate roles: the overseer drives the process, escalates, and evidences closure, while the accountable platform tower executes the change. Collapsing the two removes the independence that makes the closure evidence credible. Second, independent testing must be organizationally separate from the management of the target systems, a principle established in penetration testing guidance generally and reflected in the requirement for an independent testing agent in recognized control frameworks. A test performed by the team that configured the system validates the configuration, not the control.

Appendix A:
Fourteen Questions for Your Next Mainframe Security Review

These questions are designed to be asked by a security leader without deep platform expertise and answered by the platform team with evidence rather than assurance. An answer of "we would need to check" is itself a finding.

#	Question	What a strong answer contains
1	What is the shortest path from an ordinary corporate laptop to our claims master file, and when was it last measured?	A specific path, a date, and the name of who measured it
2	How many user IDs currently hold SPECIAL, OPERATIONS or AUDITOR authority, and how many are time-bound?	An exact count, a trend, and a time-bound percentage
3	How many privileged IDs have not authenticated interactively in the last 90 days?	A count with disposition per account
4	Which privileged mainframe roles are exempt from phishing-resistant MFA, and under what documented exception?	An exception register with owner and expiry
5	Do we hold an approved ESM security baseline, and what is our current measured drift from it?	A versioned baseline and a current drift report
6	How do federated identities from Active Directory or our identity provider map to mainframe user IDs, and who reviews those mappings?	A documented mapping with a named reviewer and cadence
7	Which service, started-task and batch IDs hold surrogate authority, and what business process requires it?	A per-ID justification, not a category justification
8	When was the last penetration test of the mainframe environment, who performed it, and were they organizationally independent of the team that manages it?	A date, a party, and an independence statement
9	Are backup, replica and test copies of regulated datasets in scope for our mainframe security assessment?	An explicit yes with the inventory to prove it
10	If a medium-severity finding is deferred, do we know how many attack paths it participates in?	Path-level analysis, not severity-level triage

#	Question	What a strong answer contains
11	Who is accountable for driving mainframe remediation to closure, and who verifies that closure independently?	Two different named roles
12	Could we produce, within five business days, the mainframe vulnerability evidence a regulator or cyber underwriter would ask for?	A named artifact and a recent example
13	Do we know, without checking, which reinsurers, brokers, MGAs, TPAs and bureau services can access our mainframe-resident data, and when that access was last reviewed?	A current inventory and a recent review date
14	How quickly could we prove that emergency privilege grants (catastrophe response access, conversion weekend access) were revoked when the event ended?	Revocation evidence tied to a specific event

Table 11. Fourteen questions. Use them to establish a baseline before commissioning any assessment.

Appendix B: Where the Exposure Sits in an Insurance Mainframe Estate

The eight vectors in Section 3 describe how AI changes attacker capability. The table below maps that capability onto the workloads where insurance mainframe exposure actually concentrates, so that a security leader can connect a vector to a specific business consequence without first translating platform terminology.

Workload	Contents	Impact of Compromise
Policy Administration	Policy masters, rating and underwriting data	Full book access
Claims	Claims files, reserves and adjuster notes	Fraud and payment exposure
Billing & Commissions	Banking and commission data	Financial theft
Reinsurance & Reporting	Treaty and bordereaux data	Regulatory and competitive risk
Actuarial & Analytics	Full-book extracts	Production data operating outside platform controls

Table 12. Where insurance mainframe exposure concentrates, by workload.

About This Paper

This paper was prepared by Ensono for discussion with security leadership at property and casualty, life, annuity and specialty carriers. It describes the threat environment as reflected in the public record as of 4 August 2026 and the assessment discipline Ensono applies through its Mainframe AI Risk & Security Vulnerability Assessment and its Infrastructure AI Security Vulnerability Scanning & Management managed service. Every statistic and regulatory reference in this document is footnoted to a primary or clearly authoritative source. Where the public record does not support a claim, the paper says so explicitly rather than filling the gap.

Source List

Full source detail appears in the footnotes throughout this document. The primary sources relied upon are listed below.

1. CrowdStrike, "2026 Global Threat Report" key findings: https://www.crowdstrike.com/en-us/blog/crowdstrike-2026-global-threat-report-findings/
2. CyberScoop, "Attack breakout time keeps shrinking, CrowdStrike annual report finds": https://cyberscoop.com/crowdstrike-annual-global-threat-report-attack-breakout-time/
3. Anthropic, "Disrupting the first reported AI-orchestrated cyber espionage campaign" (GTG-1002): https://www-cdn.anthropic.com/d7dd50dd1185f59be051b307150d877f2b82bd2c.pdf
4. Google Threat Intelligence Group, "Advances in Threat Actor Usage of AI Tools": https://cloud.google.com/blog/topics/threat-intelligence/threat-actor-usage-of-ai-tools
5. OpenAI, "Disrupting malicious uses of AI: October 2025": https://openai.com/global-affairs/disrupting-malicious-uses-of-ai-october-2025/
6. OpenAI, "Disrupting malicious uses of AI" (February 2026): https://openai.com/index/disrupting-malicious-ai-uses/
7. Rapid7, "Project Glasswing: What Security Leaders Should Know and Do Now": https://www.rapid7.com/blog/post/ai-what-project-glasswing-means-for-security-leaders/
8. UK AI Security Institute, "Our evaluation of Claude Mythos Preview's cyber capabilities": https://www.aisi.gov.uk/blog/our-evaluation-of-claude-mythos-previews-cyber-capabilities
9. Ireland National Cyber Security Centre, statement on Anthropic's Mythos Preview model: https://www.gov.ie/en/department-of-justice-home-affairs-and-migration/press-releases/national-cyber-security-centre-ncsc-statement-on-anthropics-mythos-preview-model-for-defensive-cyber-security-purposes/
10. Politico, "Anthropic's AI models hacked 3 organizations during testing": https://www.politico.com/news/2026/07/30/anthropic-ai-rogue-hacks-01018741
11. BMC, "BMC Survey Confirms Mainframe Platform Innovation" (2026 Mainframe Survey): https://www.bmc.com/newsroom/releases/bmc-survey-confirms-mainframe-platform-innovation.html
12. BMC, "Mainframe Survey Financial Services: Key Takeaways": https://www.bmc.com/blogs/mainframe-survey-financial-services-key-takeaways/
13. Planet Mainframe / Rocket Software modernization survey findings: https://planetmainframe.com/2026/03/new-survey-findings-new-flashsystem-portfolio-and-more/
14. Arcati, "Mainframe Navigator 2025 User Survey": https://planetmainframe.com/wp-content/uploads/arcati/Arcati-Mainframe-Navigator-2025-User-Survey.pdf
15. Broadcom, "Modernize Security with ACF2 and Top-Secret Version 17.0": https://news.broadcom.com/mainframe-software/modernize-security-with-acf2-and-top-secret-version-17-0
16. Kaspersky Securelist, "Deconstructing RACF in z/OS and uncovering security issues": https://securelist.com/zos-mainframe-pentesting-resource-access-control-facility/116873/

17. Guide Share Europe UK, mainframe External Security Manager vulnerability taxonomy (2020): https://conferences.gse.org.uk/2020/presentations/1AZ.pdf
18. IBM, "Cost of a Data Breach Report 2025": https://www.ibm.com/reports/data-breach
19. IBM Think, "2025 Cost of a Data Breach: navigating AI": https://www.ibm.com/think/x-force/2025-cost-of-a-data-breach-navigating-ai
20. NAIC, "Model Bulletin on the Use of Artificial Intelligence Systems by Insurers": https://content.naic.org/sites/default/files/cmte-h-big-data-artificial-intelligence-wg-map-ai-model-bulletin.pdf
21. NAIC, artificial intelligence topic page: https://content.naic.org/insurance-topics/artificial-intelligence
22. Quarles & Brady, state adoption tracker for the NAIC AI Model Bulletin: https://www.quarles.com/newsroom/publications/nearly-half-of-states-have-now-adopted-naic-model-bulletin-on-insurers-use-of-ai
23. Mayer Brown, "NYDFS Issues Two Industry Letters on Responding to a Heightened Cybersecurity Threat Environment, Including Frontier AI Models": https://www.mayerbrown.com/en/insights/publications/2026/05/nydfs-issues-two-industry-letters-on-responding-to-a-heightened-cybersecurity-threat-environment-including-frontier-ai-models
24. Greenberg Traurig, "NYDFS Final Cybersecurity Rules: MFA, Asset Inventory, and Third-Party Risk": https://www.gtlaw.com/en/insights/2025/11/nydfs-final-cybersecurity-rules-mfa-asset-inventory-and-third-party-risk
25. EIOPA, "Digital Operational Resilience Act (DORA)": https://www.eiopa.europa.eu/digital-operational-resilience-act-dora_en
26. NAIC, Insurance Data Security Model Law (MDL-668): https://content.naic.org/sites/default/files/model-law-668.pdf
27. SecurityWeek, "CISA Directs Federal Agencies to Prioritize Security Patches Based on Risk" (BOD 26-04): https://www.securityweek.com/cisa-directs-federal-agencies-to-prioritize-security-patches-based-on-risk/
28. CyberScoop, "CISA directive orders agencies to prioritize vulnerability remediation" (BOD 26-04): https://cyberscoop.com/cisa-vulnerability-remediation-directive-bod-26-04/
29. FedRAMP Public Notice 0014, "FedRAMP Response to CISA BOD 26-04": https://www.fedramp.gov/notices/0014/
30. The White House, Executive Order 14409, "Promoting Advanced Artificial Intelligence Innovation and Security": https://www.whitehouse.gov/presidential-actions/2026/06/promoting-advanced-artificial-intelligence-innovation-and-security/
31. Federal Register, EO 14409 publication record (91 FR 34565): https://www.govinfo.gov/app/details/FR-2026-06-05/2026-11415
32. Congressional Research Service, summary of EO 14409: https://www.congress.gov/crs-product/IF13268
33. CISA, NSA and Five Eyes partners, "Careful Adoption of Agentic AI Services": https://www.cisa.gov/resources-tools/resources/careful-adoption-agentic-ai-services
34. Crowell & Moring, "American and Allied Cyber Agencies Issue First Joint Guidance on Securing Agentic AI": https://www.crowell.com/en/insights/client-alerts/american-and-allied-cyber-agencies-issue-first-joint-guidance-on-securing-agentic-ai
35. Rapid7, "Formalizing Red Teaming Offensive Methodology as a Multi-Agent AI Architecture": https://www.rapid7.com/blog/post/so-red-teaming-offensive-methodology-multi-agent-ai-architecture/