Global Report on the Current State of Cybersecurity and AI Governance (2025)

A report by the Cybersecurity & AI Governance Initiative (CAGI)

Global Report on the Current State of Cybersecurity and AI Governance (2025)

Introduction

The dual rise of digital threats and artificial intelligence (AI) is reshaping how organizations approach governance and risk management. Cyber-attacks are growing in sophistication and impact, from state-sponsored hacks on critical infrastructure to criminal exploits of software supply chains, even as AI technologies become embedded in business and society. This convergence brings tremendous opportunity alongside unprecedented risks. It has become evident that strong governance is the linchpin for ensuring security and trust in the digital age. Leaders across industries are recognizing that cybersecurity and AI governance are no longer technical afterthoughts, but board-level imperatives. The challenge is to govern proactively, with foresight rather than mere compliance, so that innovation can flourish safely.

This report provides a comprehensive global overview of the current state of cybersecurity and AI governance in 2025, analysing major trends, risks, and governance failures, and highlighting gaps between today's practices and tomorrow's demands. It distils insights from international frameworks (NIST, ISO, etc.), the perspective of the CAGI (Cybersecurity and AI Governance and leading research.

Finally, it outlines 10 actionable improvements for organizations to strengthen their AI governance (with real-world examples), proposes an AI Governance Framework Template covering key pillars (risk oversight, accountability, traceability, ethics, resilience), and offers a strategic outlook on future-proofing governance for emerging challenges like quantum threats, autonomous AI, and cross-border complexities.

The core message, echoing CAGI's ethos, is that effective governance should serve as forward-looking guidance, a means of anticipating and mitigating risks, rather than a brake on innovation. By aligning with global standards and integrating efforts across cybersecurity, AI, and even quantum readiness, institutions can transform governance from a reactive necessity into a strategic enabler of trust, resilience, and responsible progress.

CAGI

1. Major Trends, Risks, and Failures in Cybersecurity and AI Governance

1.1 Escalating Cyber Threats and Governance Challenges

The cyber threat landscape in 2025 is more perilous and complex than ever. Organizations face relentless waves of ransomware, data breaches, and state-sponsored attacks targeting critical infrastructure. Notably, some threat actors have attempted to cause physical harm via cyber mean, for instance, a state-linked cyberattack on an industrial facility sought to trigger lethal accidents (fortunately failing, but causing $300 million in downtime). Such incidents underscore that cyber risks can transcend financial damage to endanger lives, elevating cybersecurity to a national security concern. A troubling asymmetry persists: attackers continue to find it "cheaper, faster, and easier to attack than to defend" in cyberspace. This is due in part to fragmented defences, different sectors (finance, energy, cloud providers, etc.) protect their own silos, but lack "connective tissue" for collective defence. The result is an economic imbalance where the costs of breaches are borne broadly (averaging $4.45 million per incident globally) while investments in prevention lag.

Compounding this, supply chain and software ecosystem vulnerabilities have emerged as systemic risks. A single compromised vendor or an open-source library flaw can cascade to thousands of organizations. High-profile failures like the *SolarWinds* supply chain breach and the *Log4j* vulnerability demonstrated how weaknesses in third-party components or processes can lead to governance failures with far-reaching impact. Many organizations discovered that they lacked visibility and controls over these interdependencies, a governance blind spot. Modern digital systems are deeply interdependent, and without coordinated safeguards and shared standards, "local disruptions propagate through complex, code-dependent ecosystems". This has prompted calls for more holistic governance approaches that treat cybersecurity as a collective responsibility rather than a purely individual enterprise concern.

Importantly, governance failures have often been failures of foresight and leadership rather than technology alone. Analyses of major breaches frequently reveal missed warnings, unimplemented patches, or poor risk culture as root causes, essentially, human governance lapses. For example, the Equifax breach (2017) was traced in part to an unpatched known vulnerability and insufficient internal accountability for data protection, a classic case where compliance checklists existed on paper but effective governance did not permeate organizational behaviour.

Similarly, the 2021 Colonial Pipeline ransomware incident highlighted gaps in contingency planning and network segregation, governance issues that allowed an IT breach to disrupt critical OT (operational technology) operations. These cases illustrate that technical defences cannot compensate for weak governance practices.

As the World Economic Forum (WEF) observed, "cybersecurity today is 80% governance and 20% technology", meaning success depends largely on leadership, policies, and culture rather than any single security product.

Key governance challenges include aligning incentives and accountability in organizations. Often, the cost of cybersecurity failures is borne by society at large (through systemic impacts), whereas the cost of prevention rests with individual firms, leading to underinvestment. Short-term business pressures can overshadow long-term security, and boards may struggle to quantify cyber risks in business terms. Moreover, executive oversight has historically treated cybersecurity as a technical silo rather than a strategic enterprise risk. This is changing, regulatory scrutiny and high-profile incidents are forcing boards and CEOs to pay attention, but many leaders still face a "literacy gap" in asking the right questions about risk appetite, resilience, and dependencies. The lack of a common baseline for what constitutes "sufficient" cybersecurity also contributes to inconsistent governance; in a WEF survey, 72% of organizations said their cyber risk increased in the past year, yet private-sector leaders diverged on the standards and measures needed for adequate security. This "governance void" leaves boards unsure how to translate regulatory guidance into concrete risk management actions, highlighting the need for clearer frameworks and shared metrics.

1.2 The Double-Edged Sword of AI in Cybersecurity

AI has rapidly become a strategic factor in both cyber offense and defence, introducing new governance considerations. Malicious actors are leveraging AI to supercharge cyber attacks, automating tasks that previously required human effort. According to the WEF's Global Cybersecurity Outlook 2024, experts anticipate AI will tilt the advantage to attackers in the near term, as tools like generative models enable more convincing phishing campaigns, faster malware development, and the creation of hyper-realistic deepfakes for social engineering. For example, generative AI can craft personalized scam emails at scale, defeating traditional spam filters and fooling victims through context-aware language. We have already seen "deepfake" audio used in fraud (impersonating CEOs' voices) and AI-assisted reconnaissance that maps out targets' networks far quicker than a human could. Security analysts warn that large language models (LLMs) could help less-skilled hackers generate sophisticated exploits, lowering the entry barrier for cybercrime.

In essence, AI is turbocharging the volume and sophistication of threats, forcing defenders to contend with attacks at machine speed.

On the flip side, AI offers powerful tools for cybersecurity defence, but not a silver bullet. Organizations are deploying AI for threat detection, anomaly spotting, and automated incident response. For instance, machine learning models can sift through logs to identify patterns of attack or use predictive analytics to anticipate which vulnerabilities are most likely to be exploited. The WEF notes that while LLMs and AI "can be harnessed for tasks like classifying data or automating threat searches," they "cannot replace the creative and nuanced human element…essential for combating threats". In other words, AI can augment security teams by handling data-intensive monitoring and first-line analysis, but human expertise is still critical for complex decision-making and threat hunting. An emerging best practice is to use AI as a force multiplier, e.g. employing AI 'co-pilots' to triage alerts or suggest responses, while maintaining human oversight to avoid false positives or strategic blind spots. A governance challenge here is ensuring the AI tools themselves are reliable, transparent, and secure (to avoid scenarios where attackers manipulate defensive AI or where the AI misses novel attack techniques due to training bias).

AI governance failures and incidents have already provided cautionary tales. A Boston Consulting Group (BCG) study found AI-related incidents (ranging from algorithmic faults to unintended behaviours) rose 21% in the past year, underlining that these risks are no longer theoretical. One notorious example was the case of Zillow's house-price prediction AI, which in 2021 was used to algorithmically buy and sell homes. The model began to severely misjudge prices in a changing market, leading to massive financial losses and the shutdown of Zillow's entire iBuying business. In hindsight, critics noted governance lapses: the company relied too heavily on the algorithm without robust human review or contingency plans for when the model went out-of-distribution (e.g., during rapid market shifts).

This illustrates the need for model risk management and human-in-the-loop oversight in AI deployments, especially in high-stakes domains. Another infamous incident was Microsoft's "Tay" chatbot in 2016, which was released to Twitter with minimal content moderation controls; within hours, trolls exploited it to spew racist and inflammatory tweets, forcing Microsoft to shut it down. The Tay debacle is a textbook case of governance failure in AI ethics and content oversight, the AI system was not properly sandboxed or aligned with company values, and there was no mechanism to prevent or detect misuse until it was too late.

In the public sector, the Dutch childcare benefits scandal (2013–2019) exemplified how unchecked AI can cause systemic injustice. Dutch tax authorities used an algorithmic risk model to flag fraudulent benefit claims, but the model was biased and insufficiently governed, falsely accusing thousands of innocent families (often of immigrant background) of fraud. The result was devastating: wrongful penalties, family hardships, and eventually a national scandal that led to the government's resignation. A later investigation concluded that "this scandal was not caused by 'bad AI' but by irresponsible design, governance, and oversight", officials implemented an algorithm without sufficient transparency, appeals processes, or ethical review. The lesson is clear: AI systems, especially in sensitive

areas, require rigorous governance including bias audits, accountability for outcomes, and avenues for human intervention. When such governance is absent, AI can inflict real harm on society and erode trust in institutions.

1.3 International Frameworks Shaping Cyber and AI Governance

In response to these challenges, international standards and frameworks are evolving to guide organizations in managing cybersecurity and AI risks. Two of the most influential sources of guidance are the US NIST (National Institute of Standards and Technology) frameworks and the ISO (International Organization for Standardization) standards, which are increasingly referenced worldwide as best practices.

- **NIST Cybersecurity Framework (CSF):** First released in 2014 (with global uptake across industries), the NIST CSF provides a structured approach to cybersecurity risk management through five core functions, Identify, Protect, Detect, Respond, Recover. In 2023, NIST released CSF 2.0 (draft), marking the first major update. Notably, CSF 2.0 introduces an expanded scope and a sixth core function: "Govern". This added Govern function underscores that effective cybersecurity requires oversight and alignment from the top. The updated framework explicitly calls for integrating cybersecurity into enterprise governance, establishing roles, policies, communication flows, and continuous improvement mechanisms.

  By elevating "Govern" alongside technical functions, NIST reflects a broad trend: treating cybersecurity as an ongoing governance process rather than a one-time technical fix. CSF 2.0 also emphasizes supply chain risk management and cybersecurity measurements, which help organizations address some of the systemic issues noted earlier.

- **NIST AI Risk Management Framework (AI RMF):** Released in January 2023, NIST's AI RMF 1.0 is a voluntary framework specifically for AI governance. It offers principles and processes to incorporate trustworthiness into the design, development, use, and evaluation of AI systems. Built through a multi-stakeholder process, the AI RMF outlines core functions like Map, Measure, Manage, and Govern AI risks. It aligns with attributes of trustworthy AI (such as validity, fairness, accountability, transparency) and provides guidance on how organizations can assess and mitigate AI-specific risks (from bias to security vulnerabilities). The AI RMF is designed to complement sector-specific regulations and other standards, for instance, it can be cross-walked to ISO and the EU AI Act requirements. NIST has also published an AI RMF "Playbook" and profiles (including a 2024 profile for Generative AI risk management) to help operationalize these guidelines.

  The emergence of NIST's AI framework signifies a maturing consensus that AI needs a governance framework analogous to cybersecurity, flexible, risk-based, and continually updated as AI technology evolves.

- **ISO/IEC 27001 and Cybersecurity Standards:** ISO 27001 is a well-established international standard for Information Security Management Systems (ISMS). Many organizations globally are certified to ISO 27001, which prescribes a risk management process and a set of security controls covering areas like asset management, access control, cryptography, incident response, and compliance. Adhering to ISO 27001 helps organizations structure their cybersecurity governance, ensuring they have the fundamental policies and processes in place. Additionally, ISO has issued related standards (e.g., ISO 27002 as a control implementation guide, ISO 27005 for risk assessment, and ISO 27701 for privacy information management) that extend governance into specialized domains.

  These standards provide internationally recognized benchmarks, aligning with them can both improve security posture and demonstrate due diligence to partners and regulators. Alignment with such standards is a key pillar of CAGI's approach, which stresses using global best practices as a baseline for governance (rather than reinventing the wheel in isolation).

- ISO/IEC 42001:2023, AI Management System Standard: Recognizing the need for AI-specific governance, ISO recently introduced ISO/IEC 42001 in late 2023, a new standard for Artificial Intelligence Management Systems (AIMS). Dubbed the "global standard for AI governance," ISO 42001 provides a structured framework for organizations to implement responsible AI governance aligned with international best practices. It covers requirements such as risk management for AI (identifying and mitigating biases, safety risks, privacy issues), AI lifecycle management (from design to deployment to monitoring), and oversight of third-party AI components. The standard promotes ethical AI principles, transparency, fairness, accountability, and continuous monitoring and improvement of AI systems.

  By obtaining ISO 42001 certification, companies can demonstrate that they have instituted governance processes to build trustworthy AI, which is increasingly important as regulations like the EU AI Act loom. For example, ISO 42001 compliance can help meet upcoming EU requirements on AI risk and documentation, much as ISO 27001 helps with GDPR compliance.

  This standard is a timely tool for governance: as one industry expert noted, it "helps organizations build trust, achieve AI compliance, and align with international best practices" for responsible AI. Early adopters, such as some Swiss firms, view ISO 42001 as a cornerstone to prepare for stricter AI regulations while maintaining innovation.

- ISO/IEC 38507:2022, Governance of AI for Organizations: Another notable standard is ISO 38507, which provides guidance for the governing bodies (e.g., boards) of organizations on the implications of AI. It essentially extends the principles of corporate IT governance (from ISO 38500) to the AI context, advising boards on how to ensure AI use is aligned with the organization's objectives, values, and legal obligations. This includes establishing clear roles and responsibilities for AI oversight, integrating AI risks into enterprise risk management, and fostering a culture of accountability and ethics around AI. ISO 38507 is about educating leadership on AI governance, a critical need given many boards are still tech-illiterate when it comes to AI. By following such guidance, boards can ask smarter questions (e.g., "What bias mitigation do we have in place for our AI models?") and set the tone for responsible AI from the top.

- Other International and National Frameworks: Beyond NIST and ISO, there are numerous frameworks shaping the governance landscape. The OECD AI Principles (2019), endorsed by 42 countries, articulate values like fairness, transparency, and robustness for AI and have spurred national AI policies.

  The European Union's AI Act (expected to take effect around 2025–2026) will impose legally binding requirements on "high-risk" AI systems (e.g., in healthcare, finance, employment), organizations will need governance processes for risk assessment, documentation, human oversight, etc., to comply. In cybersecurity, the EU's NIS2 Directive (2023) updates network and information security requirements across critical sectors and supply chains, effectively raising the bar for governance (e.g., mandating executive accountability and risk management measures for a wider range of companies). Industry-specific frameworks also contribute: e.g., the automotive sector's ISO21434 and UNECE WP.29 regulations for vehicle cybersecurity, or healthcare's HIPAA Security Rule for protecting health data, each requiring governance controls in those contexts.

In summary, international frameworks provide a blueprint for governance, and alignment with them is both a best practice and increasingly a expectation from regulators and partners. CAGI's position strongly advocates "alignment with international standards" as a cornerstone of governance programs, to ensure consistency, interoperability, and credibility. By leveraging standards like NIST and ISO, organizations can avoid governance blind spots and keep pace with the state of the art.

However, frameworks alone are not a panacea, their effectiveness depends on how earnestly organizations implement them. A recurring issue has been the "paper compliance" syndrome, where firms obtain certificates or publish policies but fail to live the principles in daily operations. Thus, the focus is shifting to operationalizing

these frameworks, for instance, using continuous controls monitoring, audits, and cultural change to make sure that what's on paper translates to real secure and ethical practices.

Successful governance thus marries the guidance of frameworks with strong organizational commitment.

1.4 Notable Governance Failures and Lessons Learned

Despite the available guidance, gaps in governance have led to high-profile failures from which important lessons can be drawn:

- Inadequate Risk Oversight: A common theme is boards or executives not actively overseeing technology deployments. The Boeing 737 Max crisis (2018–2019), while an aircraft safety example, is instructive: automated MCAS software introduced risks that were not fully understood or governed, contributing to tragic accidents. Investigations found that business pressures and siloed engineering decisions overrode sound risk governance. This underscores that whether it's flight control software or an AI system in a hospital, governance must ensure safety and ethics are never subordinated to speed-to-market or cost savings.

- Bias and Ethics Failures: Aside from the Dutch benefits scandal noted, other AI ethics failures include COMPAS recidivism algorithm bias (where a judicial AI tool was found to disproportionately flag black defendants as higher risk, raising alarms about transparency and fairness), and hiring algorithms that discriminated (e.g., Amazon's 2018 recruiting AI that was abandoned for penalizing female candidates, reflecting the biases in its training data). These cases show that without proactive governance to audit data and outcomes for bias, AI can perpetuate or even amplify discrimination. They have led to calls for routine AI ethical impact assessments and external audits as part of governance.

- Data Governance Lapses: Several incidents highlight poor data controls: the Cambridge Analytica scandal (2018) revealed how Facebook's lax data governance allowed a third-party to harvest millions of user profiles and deploy algorithms to influence elections, causing public outrage and regulatory action. More recently, a class-action lawsuit against a media company (Paramount, 2023) exposed that sharing subscriber data with AI analytics vendors without proper consent or safeguards can result in legal liability and reputational damage. It exemplifies how governance must cover not just AI models but the data pipeline feeding AI, ensuring privacy and compliance are maintained.
- Security-Privacy Trade-off Missteps: Some failures come from not anticipating how AI and cyber intersect. For example, in 2022 it emerged that an AI-powered customer support chatbot at a large bank was exploited to reveal private customer information, attackers probed it with cleverly crafted inputs, exploiting the AI's training data. The bank had rolled out the chatbot for efficiency but had not fully threat-modelled how it could be abused (a governance miss). This scenario taught many that AI systems exposed to users (like chatbots) need rigorous security evaluation (adversarial testing, prompt injection defences, rate limiting, etc.) just as any other part of the IT system.

- Lack of Incident Preparedness: Governance also means preparing for failures. Some organizations have suffered worse impacts from incidents simply because they lacked an incident response plan or clarity of roles when AI or IT incidents occurred. For instance, a global shipping company hit by malware found that its lack of a coordinated response plan led to a chaotic reaction, compounding losses. In the AI realm, if a company's AI system produces a harmful outcome (say a medical AI misdiagnoses patients), does the organization have a response plan to halt the AI, notify affected parties, and remediate? Many do not, a gap that needs closing.

The common lesson across these examples is that technology will fail in unexpected ways, and only robust governance can mitigate the fallout. Whether it's a faulty algorithm, a malicious intrusion, or a misuse of data,

governance structures, clear accountability, foresight in risk assessment, ethical principles, and practiced response protocols, differentiate organizations that navigate the issue with minimal damage from those that spiral into crisis. As such, examining past failures fuels improvements in frameworks and practices going forward. For instance, incidents of AI bias are driving inclusion of fairness and bias checks in model development standards, while cyber incidents in supply chains are accelerating adoption of zero-trust architectures and third-party risk governance. Each failure has become fodder for new governance tools, essentially, governance is evolving iteratively, often one headline at a time.

2. Gaps Between Current Practices and Future Demands

Despite growing awareness, current cybersecurity and AI governance practices often lag behind what the future demands. The landscape is evolving so rapidly that many organizations find themselves governing yesterday's risks, while tomorrow's are emerging unaddressed. Here we provide a comparative overview of key gaps:

- From Reactive to Proactive (Foresight):

  Today's practice: Many governance programs are still largely reactive or compliance-driven. Firms patch systems after a breach, or address AI ethics only once negative press erupts. Security investments are often justified by last year's incidents, and AI oversight kicks in post deployment (if at all). Future demand: Governance must be anticipatory and preventive. This means employing foresight techniques, *threat modelling, scenario planning, "red teaming" AI systems*, to predict and mitigate risks before they materialize. CAGI emphasizes "governance as foresight, not restriction," meaning leaders should proactively set guardrails informed by imagining future threat scenarios, rather than waiting to simply restrict technologies after damage is done. For example, rather than forbidding use of generative AI outright, a forward-looking approach establishes guidelines for safe use (data handling rules, human review for AI outputs, etc.) so innovation can continue with risk mitigated.

  The gap: currently, few organizations have formal processes to do horizon-scanning for emerging risks like quantum or autonomous AI misbehaviour. Closing this gap may involve creating foresight roles or committees in governance structures (similar to how some boards now have technology and innovation subcommittees).

- From Siloed to Integrated Governance:

  Today's practice: Cybersecurity, AI governance, data governance, and IT governance often operate in silos. One team might handle infosec, another focuses on AI ethics or compliance, rarely coordinating their efforts. Quantum computing risks, if considered at all, might be left to an R&D function, separate from core security governance.

  Future demand: Integration across domains. The next decade's challenges (e.g., an AI-driven cyberattack, or a quantum-enabled breach) will span multiple domains, so governance must be unified. Cybersecurity, AI, data, and even quantum risk management need to converge into a cohesive governance framework. This means establishing cross-functional governance bodies that include cyber, AI, privacy, risk, and business representatives. It also means using common risk taxonomies and dashboards so that, for instance, a data privacy issue in an AI system is seen not just as a compliance issue but also as a security and ethical issue. CAGI's core messaging stresses "integration across cybersecurity, AI, and quantum readiness", in practice, this could be a single governance framework or steering committee that oversees digital risk holistically.

  The gap: currently, integration is more the exception than the norm. Many organizations lack mechanisms for, say, the CISO and the Head of AI to jointly evaluate an AI system's security; or for the risk committee to understand technical AI failures. Bridging this gap may involve restructuring (e.g., creating a Chief Trust

Officer or similar role responsible for end-to-end digital trust) or at least regular joint workshops between siloed teams.

- From Compliance-Based to Risk-Focused and Ethical Governance:

  Today's practice: A lot of governance today is checkbox-driven, meeting regulatory requirements (GDPR, SOX, industry standards) or obtaining certifications. While necessary, this can create a false sense of security. Organizations might be "compliant" but not actually secure or responsible. For AI, since hard regulations are only just emerging, many firms simply have high-level ethical principles on their website, but no concrete processes to enforce them.

  Future demand: A shift to outcome-focused governance that genuinely reduces risk and enforces ethics, even beyond what regulations explicitly require. This involves adopting enterprise risk management for cyber and AI, quantifying and tracking risk reduction over time, not just ticking controls. It also means operationalizing ethics: for example, instead of just stating "we value fairness," an organization would implement bias testing in model development, diverse review panels, and a process to handle ethical dilemmas or external complaints about AI decisions. Future governance will likely be judged on metrics like reduction in incidents, stakeholder trust levels, and resilience demonstrated in crises, rather than on audit scores alone.

  The gap: currently, few organizations measure governance success in terms of risk outcomes. There is a need to develop new KPIs (e.g., "time to detect and contain an AI failure" or "percentage of projects reviewed for ethical risks pre-launch") to drive meaningful governance improvements.

- From Technical Isolation to Board & Executive Engagement:

  Today's practice: Governance of technology has often been relegated to IT departments or mid-level compliance officers. Boards get involved mainly after something goes wrong. In many companies, cybersecurity is still not a regular board agenda item, and AI governance may be virtually absent at the board level unless the company's core product is AI.

  Future demand: Active board oversight and C-suite leadership on tech governance. As WEF notes, breaches and AI failures can "destabilize societies… security becomes a question of governance, and of leadership". Executives need sufficient literacy to interrogate management: e.g., "What's our risk if our AI makes a wrong decision? How do we recover? Are we within appetite?". Some regulators are pushing this direction (for instance, the U.S. SEC now requires public companies to disclose cyber governance practices at the board level, and the EU AI Act may require a designated "AI compliance officer" for high-risk AI). The expectation is that leaders treat digital risks on par with financial or legal risks.

  The gap: many current leaders did not come up in a world where AI and cyber were core concerns, so there's a skills and mindset gap. A survey in Accenture's 2025 report found only 36% of tech leaders recognized that AI is outpacing their cybersecurity capabilities, suggesting a large portion may be underestimating the risk. Bridging this gap involves education (upskilling boards on tech topics), possibly recruiting board members with cyber/AI expertise, and instilling a culture where asking tough questions about technology is the norm. It might also involve governance reforms like establishing dedicated board subcommittees for Technology or Trust (similar to audit or risk committees).

- From Point-in-Time to Continuous and Adaptive Governance:

  Today's practice: Governance activities often happen periodically, e.g., annual risk assessments, quarterly policy reviews, one-time model validation before deployment. The fast-changing threat and technology

environment means such static approaches leave long gaps during which new vulnerabilities or changes in AI behaviour might go unnoticed.

Future demand: Continuous monitoring and adaptive governance. For cybersecurity, this means real-time visibility into control effectiveness (for example, continuous control monitoring tools, ongoing threat intelligence feeds that inform risk posture adjustments week by week). For AI, it means continuously monitoring models in production for drift or anomalies (e.g., an uptick in prediction errors or a shift in input data characteristics) and having processes to recalibrate or pull back models promptly. Adaptive governance also entails agile policy-making, updating policies as new scenarios emerge (e.g., quickly formulating a policy on employee use of a new generative AI tool rather than waiting months).

The gap: currently, many organizations lack the infrastructure or processes for this. They might not be logging the right data to notice subtle issues, or their governance committees meet too infrequently to keep up with new developments. Moving to a near-real-time governance stance may require investing in automation (for instance, tools that automatically scan AI models for bias every time they are retrained, or security posture dashboards that update daily). It also requires an organizational willingness to iterate on governance documents and decisions frequently, something bureaucratic cultures may find challenging.

- From Narrow Metrics to Holistic Resilience Indicators:

  Today's practice: Success in cybersecurity is often measured in technical terms, number of incidents, compliance scores, etc. AI success might be measured by accuracy of models or revenue from AI initiatives, with little integration of risk.

  Future demand: A focus on resilience and trust as key outcomes. This means developing indicators that reflect an organization's ability to withstand and bounce back from digital disruptions. For instance, "mean time to recover from a cyber incident" or "percentage of AI systems with robust fallback plans" are resilience indicators. Trust indicators might include customer/stakeholder sentiment regarding the organization's handling of AI (did they communicate transparently? do users feel safe using the AI?). The alignment with international standards also provides metrics, e.g., achieving ISO certification or meeting NIST framework tiers can be milestones, but ultimately, resilience must be tested (through drills, audits, or even external stress tests).

  The gap: currently, very few organizations have such metrics. The future will likely demand it, as stakeholders (from investors to regulators) ask not just "Are you secure/compliant?" but "Can you continue operations under duress? Can we trust your AI with our data/lives?". Bridging this gap may involve incorporating scenario-based testing (like cyber war-gaming, AI failure simulation) into regular governance and reporting on the results.

In essence, the current state of governance often reflects a step or two behind the cutting edge of risk, whereas future demands call for leaping ahead. Bridging these gaps will not be easy, it requires cultural change, investment, and often, new skill sets, but it is necessary. Encouragingly, we see early adopters making strides: e.g., only 10% of organizations today are in what Accenture calls the "Reinvention-Ready" zone, demonstrating both strong security capabilities and integrated strategy.

Those organizations suffer significantly fewer successful attacks and recover faster. They serve as proof that the advanced governance practices described are not just theoretical ideals but achievable, with tangible benefits. The goal for the rest should be to close the gap, moving from exposed or merely progressing, into that resilient, forward-ready posture

3. Top 10 Actionable Improvements to Strengthen AI Governance (with Examples)

Considering the trends and gaps identified, organizations can take concrete steps to bolster their AI governance. Below are ten actionable improvements that can be pursued immediately, each supported by real-world examples or challenges that illustrate why it matters:

1. Establish Clear Executive Accountability for AI Governance, *Assign top-level ownership and create governance structures for oversight.* One of the first steps is to make AI governance an explicit responsibility at the executive and board level. This could mean appointing a Chief AI Ethics or AI Governance Officer, or expanding the mandate of an existing risk committee to cover AI. The key is that someone at the top is accountable for the safe and ethical deployment of AI. For example, Mastercard formed an AI Governance Council to oversee AI projects across the company, ensuring consistency with their standards. Likewise, several large banks (e.g., Citi, Deutsche Bank) have added AI oversight to board risk committees as they deploy AI in trading and customer service. Without clear ownership, AI initiatives can "fly under the radar," leading to incidents like the bias issues at Amazon's hiring AI, which reportedly lacked proper executive oversight.

Action: Charter a cross-functional AI governance committee (including legal, IT, ethics, business leaders) that meets regularly to review AI use cases, set policies, and report to the board. Support it by defining a RACI (Responsible, Accountable, Consulted, Informed) matrix for AI decisions.

Challenge: A common challenge is finding the right leadership, many executives admit they don't fully understand AI. Organizations may need to invest in training their leaders or hiring external experts. Nonetheless, making someone visibly accountable drives focus and resources toward governance. As the saying goes, "what interests the boss fascinates the employees", if the CEO and board make AI governance a priority, it permeates the culture.

2. Develop an AI Governance Framework and Policy Suite, *Document the principles, policies, and procedures that will govern AI use, aligned with international standards.* Many organizations today lack a coherent set of AI policies beyond perhaps an AI ethics statement. To strengthen governance, firms should codify how AI is to be developed and used. This includes policies on data quality and bias mitigation, model validation and performance thresholds, explainability requirements, privacy and security controls for AI systems, and compliance with regulations. A good approach is to align this framework with known standards like NIST's AI RMF or ISO 42001 so that it covers the full lifecycle. For instance, Telstra (an Australian telecom) implemented an internal AI governance framework that requires every new AI project to go through an "AI ethics & risk assessment" checkpoint, documenting potential impacts and mitigation before launch. By having such a framework, they ensure consistency and traceability, no AI system goes live without having followed due process.

Action: Draft an AI governance charter and associated policies, and integrate them into your existing governance documents (like adding an AI section to your IT governance manual or risk management framework). Ensure it covers the pillars of governance, from risk and accountability to ethics and resilience (the next section's template offers a starting point).

Challenge: Avoid making the framework too abstract; operationalize it with specific standards (e.g., "all customer-facing AI must achieve at least 95% accuracy and have undergone fairness testing on protected attributes"). Also, keep it agile, schedule periodic updates as technology and laws evolve. The benefit of a strong framework is evident when facing external scrutiny: when regulators or clients ask "How do you govern AI?", having a documented, standard-aligned answer builds trust.

3. Integrate AI Risk Management into Enterprise Risk Processes, *Treat AI risks as part of enterprise risk management (ERM), with regular risk assessments and reporting.* AI projects often start as innovation experiments, sometimes skunkworks-style, and may not go through the usual risk oversight. This must change. Organizations should include AI risks (ethical, operational, cyber, reputational) in their enterprise risk register and assessment processes. For example, a global insurance company recently updated its ERM to list "AI model risk"

as a top 10 enterprise risk, alongside credit risk and cybersecurity. This forces each business unit to report on how they're controlling AI risks.

Action: During quarterly risk reviews or annual ERM cycles, evaluate AI alongside other risks: What's the worst-case scenario if an AI misbehaves? What new threats (like model hacking) exist? Assign risk owners and mitigation plans. Scenario analysis is useful, e.g., simulate an AI failure scenario and gauge potential financial/legal impact. The board should see AI risk summaries regularly, just as they see financial or compliance risks.

Challenge: Quantifying AI risk is new territory. Unlike, say, credit risk, there's less historical data. But qualitative assessment can suffice initially (e.g., using a heat map for AI risks). Another challenge is that AI risks can be interdisciplinary (technical, ethical, legal combined), requiring collaboration between data scientists, risk officers, and business leaders. Overcoming silos (as per the integration goal) is necessary here.

4. Strengthen Data Governance and Quality Controls for AI, *Ensure the data feeding AI models is governed for quality, bias, security, and compliance.* Many AI failures and biases trace back to the data used. Thus, improving AI governance means tightening data governance. This includes maintaining data inventories, setting data quality standards, access controls, and provenance tracking (knowing where data comes from and how it's been processed). A real-world example: JPMorgan Chase implemented a robust data governance program as part of its AI efforts, creating a data catalogue and lineage tracking to know exactly what data is used in each model (and with what permissions). This paid off when they were able to quickly answer regulatory inquiries about their AI's training data origin and privacy protection.

Action: Leverage existing data governance structures (like a Data Governance Council) to specifically address AI needs. Mandate that any dataset used for AI training or analysis be profiled for biases and quality issues. Use tools to detect skew or under-representation in data (for instance, ensuring a facial recognition training set has diverse demographics to prevent racial bias). Also, implement data minimization and privacy techniques for AI, e.g., anonymization or synthetic data, to reduce the chance of privacy breaches.

Challenge: AI teams often want "all the data" for better performance, which can conflict with principles of minimization or with siloed data ownership. Overcoming this requires a balance, governance should facilitate access to good data for AI (to enable innovation) but under controlled conditions. Data governance policies may need updates to cover AI-specific issues, like prohibiting use of data that was collected for one purpose to train an AI for another purpose without consent (to avoid cases like the aforementioned Paramount lawsuit on data misuse).

5. Embed Security and Privacy by Design into AI Development, *Extend cybersecurity and privacy practices into the AI model development lifecycle.* AI systems introduce new attack surfaces, model parameters can be tampered with, inputs can be manipulated (adversarial examples), and models can inadvertently leak sensitive data from training (membership inference attacks). Therefore, organizations should incorporate secure development life cycle (SDLC) practices for AI, akin to what's done for traditional software. For example, Microsoft and Google have published guidance on "secure AI engineering", requiring steps like threat modelling for AI systems, testing models against adversarial inputs, and ensuring encryption of model artifacts.

Action: If your organization has a product/security development lifecycle (like many do under frameworks such as Microsoft SDL or OWASP SAMM), explicitly integrate AI. This could mean: conducting security reviews for new models, pen-testing AI APIs, and using privacy-enhancing techniques (like differential privacy) for models that train on user data. Also, treat AI models as sensitive assets, control access to model APIs, log all predictions for abuse monitoring, and enforce access controls on model training code and data.

Challenge: Many data science teams are not versed in security, and many security teams are not versed in AI. Bridging this gap is crucial, perhaps by embedding a security expert in AI projects or training data scientists on secure coding and privacy. The effort is worthwhile: consider that a survey found 77% of organizations lack

foundational AI security practices to safeguard models and data pipelines, a startling figure that attackers will exploit. Closing this gap through security-by-design can prevent incidents like the one where hackers manipulated an AI vision system by subtly altering street signs (to misguide a self-driving car). It's much harder to bolt on security later; building it in from design stage is far more effective.

6. Ensure Transparency, Explainability, and Traceability of AI Decisions, *Implement measures so that AI-driven decisions can be understood, audited, and trusted by humans.* A cornerstone of AI governance is making AI "explainable", especially for high-stakes use cases, and having a record of AI decision logic (traceability). This addresses both ethical and risk concerns: stakeholders have a right to know how an AI made a choice, and organizations need that knowledge for accountability. Many financial services firms provide a good example: by law, if they use AI for credit decisions, they must provide adverse action notices explaining factors for denial. Thus, banks have invested in "model interpretability" tools that can output the key variables influencing a decision (age, income, etc.) and their weights.

Action: Adopt interpretable model techniques where possible (e.g., use explainable AI (XAI) algorithms or post-hoc explanation tools like LIME/SHAP for complex models). Maintain documentation for each AI model, including its purpose, training data, algorithm type, and limitations, as part of an AI model registry. Such documentation, sometimes called Model Cards (an idea popularized by Google), provides traceability and transparency. Also, consider user-facing transparency: if AI is used in a customer interaction, disclose it (many companies now add notices: "This chatbot is AI-assisted").

Challenge: There is a trade-off, as more complex models (deep neural networks) are often less explainable. Some accuracy may need to be sacrificed for interpretability in critical contexts, governance should guide where that trade-off lies (e.g., it might be acceptable in a movie recommendation engine, but not in an autonomous driving system or healthcare diagnosis). Regulators are increasingly demanding explainability (the EU AI Act will require it for high-risk AI), so organizations that start now will be ahead. An example of what happens without this: in healthcare AI, a renowned hospital had an AI recommending treatments but could not explain the rationale to doctors, leading to distrust and eventual scrapping of the system. Transparency practices could have salvaged its utility by building doctor confidence.

7. Conduct Regular Audits and Stress Tests of AI Systems, *Periodically evaluate AI models for performance, bias, and robustness, including "red team" exercises.* Just as financial institutions run stress tests on algorithms and portfolios, AI systems benefit from periodic challenge. This might involve auditing a model's decisions on new data to see if it's still performing, or specifically testing it for bias or error rates on subpopulations (e.g., does a loan approval AI inadvertently favour one demographic?). It also involves robustness testing: checking how the AI handles slight distortions or attempts to fool it (adversarial inputs). For example, Facebook (Meta) created an "AI Red Team" to intentionally try to trick their AI systems and identify vulnerabilities, such as the propensity of their content filters to be bypassed by clever phrasing.

Action: Set up an internal audit schedule for AI. This could be quarterly evaluations for critical models and semi-annual for others. Use independent reviewers if possible (or even third-party auditors in some cases) to get fresh eyes. Track metrics over time (accuracy, bias, etc.). Additionally, run incident response drills for AI failures: simulate scenarios like "our AI makes a harmful recommendation, what do we do?" to ensure readiness.

Challenge: Unlike traditional audits, AI audits require technical expertise in ML; organizations may need to train internal audit teams or work with specialized firms. There's also the issue of defining thresholds, what amount of bias is "too much"? Governance bodies should set these criteria in advance (e.g., "if error rate for any group is 2x higher than others, that triggers a retraining"). By stress-testing AI, organizations can catch issues early, much like how banks use penetration testing to find cyber weaknesses before hackers do. This proactive approach might have prevented, for instance, the Zillow scenario: a thorough stress test under different market conditions could have revealed the model's brittleness before it caused huge losses.

8. Build an AI Incident Response and Monitoring Capability, *Prepare to detect, respond to, and learn from AI-related incidents or failures.* As AI becomes part of critical processes, we must be ready for when things go wrong. This means extending the notion of a Security Operations Centre (SOC) or incident response team to cover AI incidents. For example, if an AI-driven trading system starts behaving erratically, is there a kill-switch or alert? If a chatbot begins giving dangerous advice, will anyone notice before harm is done?

Action: Implement monitoring on AI systems in production. Establish baselines for normal behaviour and set up alerts for anomalies (e.g., sudden spikes in certain outputs, or significantly different input patterns that could indicate misuse). Train the incident response team on AI scenarios, perhaps create runbooks: "If AI system X produces output outside range Y, do Z." Some organizations have instituted "model emergency brakes", automated triggers that disable an AI model if predefined safe parameters are breached (for instance, if a self-driving car's vision AI starts misclassifying objects at a high rate, it hands control to a human or safe mode). Also, incorporate AI considerations into breach response plans. A data breach involving an AI (like theft of a model or poisoning of training data) may require special handling (such as retraining models, or informing users if their data influenced a flawed decision).

Challenge: AI incidents may not be as obvious as, say, a server outage. They can be subtle and prolonged (like a slow drift into biased decisions). It requires defining what constitutes an "incident" in AI, governance teams should define this. A promising practice is learning from others: industries like aviation have decades of safety incident management; similar principles (blameless post-mortems, sharing learnings across the industry) could be applied to AI. Indeed, there are initiatives for "AI incident databases" where organizations contribute anonymized reports of AI glitches and fixes, to collectively improve governance (much like air accident investigations lead to better regulations). Ultimately, being prepared for incidents turns potential crises into manageable events. An example: one hospital had an AI diagnostic tool misidentify a serious condition; because they had a review system and escalation path in place, a human doctor double-checked and caught the mistake, avoiding harm. That kind of resilience comes only with preparation.

9. Invest in Workforce Training and a Culture of Digital Responsibility, *Educate and involve employees at all levels in AI governance and cybersecurity best practices.* Governance frameworks and policies mean little if the people operating under them don't have the awareness or skills to follow them. Strengthening AI governance requires a parallel investment in human capital. This includes training technical teams (engineers, data scientists) in ethics, privacy, and security, and training non-technical teams (management, front-line staff) in understanding AI limitations and responsible use. For example, Singapore's DBS Bank embedded AI governance in its culture by training thousands of employees on an "AI ethics code of conduct" and how it relates to their roles. They coupled this with top-down messaging that everyone is a steward of data and AI trust.

Action: Roll out training modules about your AI governance policies so employees actually know them. Include case studies of AI failures and successes to illustrate abstract principles. Also, encourage an open culture where anyone can raise a concern about an AI system (maybe an internal hotline or committee to review such concerns). Often, junior data scientists might spot an issue (e.g., a bias in data) but need to feel safe to speak up.

Challenge: Time and attention are scarce; it can be hard to get non-tech staff to care about AI or get tech staff to think beyond code. One way is to tie it to personal stakes: show customer stories (positive or negative) impacted by AI, or how a cyber incident could threaten the company's success and thus everyone's jobs, making it relatable and urgent. Also, incentivize good behavior: some companies have started including "risk and ethics" components in performance reviews for relevant roles, to ensure accountability is personal. A strong culture is arguably the best defense; as the WEF noted, building "cultures of informed oversight" is critical. That means people at all levels understand the balance between innovation and risk and take initiative to do the right thing even when no one is watching. Such a culture prevented, for example, one tech company from releasing a facial recognition feature, an employee-led review raised concerns about privacy and bias, and leadership listened, refining the product before launch. Empowered, educated employees are the boots on the ground for governance.

10. Align and Integrate AI Governance with Cybersecurity and Overall GRC (Governance, Risk, Compliance) Functions, *Break down silos by embedding AI considerations into existing governance structures and vice versa.* Many organizations have established structures for IT governance, cybersecurity oversight (like a cybersecurity steering committee), and corporate compliance. Rather than reinvent the wheel, inject AI governance into these existing forums, and ensure cybersecurity is discussed in AI contexts too. For example, if there's a monthly IT risk meeting, add AI incidents or updates as an agenda item. Conversely, when AI project teams convene, have a cybersecurity representative present.

Action: Map out where decisions about technology and risk are made in your org (board committees, management committees, etc.) and consciously insert AI governance into those touchpoints. Perhaps the audit committee of the board now gets a yearly briefing on AI risk in addition to financial risk. Or the CISO includes AI systems in the scope of security audits and reports. Some companies have gone further and created joint task forces for emerging tech, bringing together legal, compliance, IT, and business to handle new challenges collectively (e.g., a "Digital Innovation Risk Committee"). The outcome is that AI is not a black box running in a corner, but part of the enterprise risk conversation and control environment.

Challenge: Turf and inertia, the IT folks might say "we handle software, AI is a data science thing" while data science might say "security will slow us down". Strong leadership needs to mandate collaboration, underlining that AI governance is not separate or optional: it's as integral as cybersecurity, which in turn is as integral as financial controls. Indeed, modern regulations treat them on similar footing. Alignment with international standards, which often cover multi-disciplinary aspects, can serve as an objective framework to rally everyone around. CAGI's message of integration resonates here, by aligning AI governance with cybersecurity frameworks (like mapping AI controls to NIST CSF categories), organizations create a unified language and avoid duplication.

The benefit is a more efficient governance system: rather than two parallel tracks (one for cyber, one for AI), there's one coherent approach, saving time and reducing gaps. In practice, this might mean when the company does its annual ISO 27001 security audit, it also assesses AI-related controls in the same audit, killing two birds with one stone and ensuring nothing slips through cracks.

By implementing these ten measures, organizations can substantially strengthen their AI governance posture. Each action not only reduces risk but also fosters innovation by building trust. For instance, if customers know your AI is well governed, that you can explain it, you secure their data, and you respond swiftly to issues, they are more likely to adopt and even embrace your AI-driven services. Thus, governance becomes a competitive advantage, not a burden.

It's worth noting that these improvements are actionable and practical; many leading organizations are already doing some or most of them, and finding that it enables them to scale AI with confidence. The journey may involve challenges and require new thinking, but the payoff is a resilient, trustworthy digital enterprise ready for the future.

4. AI Governance Framework Template, Pillars and Practices

To further aid organizations, we present an AI Governance Framework Template covering key governance pillars and their focus areas. This can serve as a checklist or foundation which can be adapted to an organization's context (e.g., industry-specific requirements). The pillars included, Risk Oversight, Accountability, Traceability, Ethics, and Resilience, reflect core aspects identified by CAGI and international standards for comprehensive governance. Table 1 below outlines each pillar with its primary focus and example practices:

Governance Pillar	Focus & Key Practices
Risk Oversight	**Board and executive-level visibility into AI and cyber risks.** Involves establishing governance bodies (e.g., AI risk committee), defining risk appetite for AI (what level of error/bias is tolerable), and integrating AI risk into enterprise risk management. Practices include regular risk reporting to the board, scenario analysis (what-if exercises for AI failures), and alignment with frameworks (NIST, ISO) to systematically identify and mitigate risks. *Example:* A board approves an AI risk appetite statement and monitors key risk indicators (like "percentage of AI models reviewed for bias before deployment").
Accountability	**Clear roles and responsibilities for AI outcomes and cybersecurity.** This pillar ensures that for every AI system or major cyber process, there is an owner accountable for its performance, ethical compliance, and risk management. It covers governance structures (who approves AI deployments, who signs off on security), and enforcement mechanisms (consequences for negligence, incentives for good governance). *Example Practices:* Designating "model owners" for AI models who must certify their models meet governance criteria, including accountability clauses in vendor contracts (so third-party AI providers are accountable too), and implementing an AI ethics code of conduct that employees sign. The tone at the top is crucial, leadership must communicate that they remain accountable for AI decisions made by the organization, even when automated.
Traceability & Transparency	**Ability to understand and audit AI decisions and data flows (sometimes called "AI governance traceability").** This pillar deals with maintaining records and documentation for AI systems: data lineage (where training data came from, how it was processed), model development history (versions, parameters, validation results), and decision logs (so any output can be traced back to inputs and model version). It also emphasizes transparency to stakeholders: providing information about how AI systems function and are governed. *Example Practices:* Using model documentation templates (model cards), enabling logging in AI applications (e.g., logging each AI-driven decision along with confidence score), and publishing summaries of AI governance efforts externally (some companies release transparency reports on AI similar to how they do for privacy). In cybersecurity, traceability means robust logging of system activities and attacks for forensic analysis. Transparency builds trust, for instance, a company might provide customers the reason why an AI denied their loan, fulfilling both compliance and trust-building.

Ethics & Fairness	**Embedding ethical principles and fairness criteria into governance.** This pillar covers proactive measures to ensure AI and cybersecurity practices uphold values like fairness, non-discrimination, privacy, and respect for user rights. It involves ethical impact assessments, bias testing, stakeholder consultation, and a process for ethical deliberation on tricky issues. *Example Practices:* Conducting an ethics review for high-impact AI deployments (perhaps via an ethics committee or external ethicist panel), implementing bias audits (checking model outcomes across different demographic groups), and ensuring user consent and privacy in data usage (aligned with regulations like GDPR). It also means having a mechanism to handle complaints or issues, e.g., if someone believes an AI decision was unfair, there is an appeal path with human review. International frameworks emphasize ethics, e.g., the OECD AI Principles and UNESCO AI Ethics Recommendation, and alignment here signals good governance. As a real example, after the Dutch scandal, the government instituted an algorithm registry and auditing process to ensure algorithms in public sector are fair and transparent.
Resilience & Security	**Ensuring systems (both AI and broader IT) are robust, secure, and able to recover from disruptions.** This pillar focuses on technical and operational preparedness: cybersecurity measures (preventing breaches or tampering of AI), robustness of AI models (resistance to adversarial attacks or unexpected inputs), and business continuity plans for digital failures. It aligns with concepts like *"secure by design"* and *"AI robustness."* *Example Practices:* Applying rigorous cybersecurity controls to AI infrastructure (access control, encryption of model files, etc.), regularly testing AI with adversarial examples to improve resilience, and having fallback plans if AI components fail (e.g., human takeover or redundant systems). It also includes incident response planning and drills, as discussed earlier, essentially the capability to absorb and quickly recover from incidents. Resilience is the metric by which governance will increasingly be judged, not just can you prevent problems, but when something inevitably goes wrong, can you limit damage and restore trust swiftly? A telling example is the difference between two companies hit by similar AI model outages, one had a backup model and communication plan, so customers were hardly affected, whereas the other was down for days and lost customers. Governance for resilience aims for the former outcome.

Table 1: Core Pillars of an AI Governance Framework and Their Key Focus Areas.

These pillars are interrelated, effective governance requires addressing all of them in an integrated way. For instance, accountability supports risk oversight (clearly assigned owners for risks), and ethics intersects with resilience (ethical lapses can cause as much crisis as technical failures). Traceability underpins oversight and accountability by providing the evidence and logs needed to monitor and enforce. When customizing a framework, organizations might add pillars (e.g., Compliance could be a distinct pillar in a heavily regulated sector, or Innovation Enablement to ensure governance also facilitates beneficial AI use), but the ones listed are considered fundamental.

Implementing such a framework often involves creating a matrix of principles vs. practices: e.g., mapping each pillar to specific controls or policies your organization will implement. Many companies start by performing a gap analysis against these pillars, checking current governance practices against each area to identify what's missing. Using the template as a guide, they can then systematically improve. For example, if the gap analysis finds no formal ethical review process exists (Ethics pillar gap), the action could be to form an AI ethics committee and require its sign-off for certain projects.

It's worth noting that adopting a governance framework does not equate to stifling innovation. On the contrary, a clear framework provides teams with guardrails within which they have freedom to innovate. Engineers and data scientists, if given guidelines on what's acceptable (data use, model types, etc.) and processes to follow, can proceed confidently rather than fear running afoul of unknown rules later. As CAGI advocates, governance should be seen as enabling foresight and trustworthy innovation, not simply as bureaucracy or restriction. A well-crafted framework achieves this balance, it is firm on critical principles (safety, ethics, accountability) but flexible on implementation (encouraging creative solutions as long as they meet the governance criteria).

In summary, the AI Governance Framework Template above offers a high-level yet actionable structure. Organizations are encouraged to adapt it to their needs, possibly expanding it into a more detailed program. Many early adopters who have taken a similar structured approach report that it helped demystify AI governance for their staff and stakeholders. It brings everyone onto the same page about what "good governance" means in concrete terms, thereby aligning efforts across departments. As we move into an era where demonstrating good governance may be required to do business (clients and regulators asking for evidence), having such a framework in place will be invaluable.

5. Strategic Outlook: Future-Proofing Governance for Emerging Challenges

Looking ahead, organizations must gear up for a future where technology challenges will be even more formidable. Quantum computing, autonomous AI agents, and global data entanglements loom on the horizon as game-changers. Governance practices that suffice today may strain or break under tomorrow's pressures. In this section, we provide a strategic outlook on how institutions can future-proof their governance, building in agility and foresight to handle what's coming in the next 5–10 years.

5.1 Quantum Threats, The Next Cryptographic Crisis

Quantum computing promises to solve complex problems beyond current computers' reach, but it also threatens to undermine the cryptographic foundations of cybersecurity. Most of today's encryption (RSA, ECC) could be cracked by a sufficiently powerful quantum computer, meaning everything from financial transactions to medical records to state secrets protected by those algorithms would be at risk. This isn't a distant sci-fi scenario, experts estimate that within a decade or less, quantum attacks could be feasible. Indeed, 73% of U.S. businesses believe "it's only a matter of time" before cybercriminals use quantum power to break today's crypto, yet paradoxically, over 80% admit they need to better evaluate their security capabilities in light of this threat. Similarly, research in Germany found 95% of organizations see quantum as a high impact risk to cryptography, but only 25% are addressing it in their risk management today. This gap between awareness and action is a governance concern.

To future-proof, governance bodies must incorporate quantum readiness into their strategies now. This includes:

- Inventory and Assessment: Organizations should inventory where they use long-lived encryption (data that needs to remain secure for years) and assess the impact if that encryption were broken. Prioritize sensitive assets, e.g., state secrets or personal data that must remain confidential for a long time (think of medical records or identities that could be misused decades later).

- **Transition Planning to Post-Quantum Cryptography (PQC):** NIST has already standardized some post-quantum algorithms (like CRYSTALS-Kyber for encryption). Governance should mandate a transition plan. This might mean requiring all new systems to use quantum-safe crypto by default by a certain date, and phasing out legacy crypto in a multi-year roadmap. Leading governments are setting such timelines (the U.S. issued a memorandum and the Quantum Computing Cybersecurity Preparedness Act, pushing agencies to migrate to PQC; Singapore's MAS issued advisories on quantum risks in 2024).
- **"Harvest Now, Decrypt Later" Mitigation:** A unique risk is adversaries stealing encrypted data today, with plans to decrypt once they have quantum capability. Thus, if an organization handles data that would still be sensitive in ~5-10 years, they should act as if it could be compromised. Mitigation could include encrypting with both classical and post-quantum algorithms (hybrid encryption) in the interim, and beefing up access controls, even if data is stolen now, making it harder to obtain in bulk buys time.
- **Quantum-safe Development and Procurement:** Future-proofing means any software developed or purchased now should be agile enough to swap out cryptographic components. Governance can set policy that all new tech must be "crypto-agile," supporting easy upgrades to new libraries. Some forward-thinking organizations even include contract clauses with vendors: requiring roadmaps for PQC support or escrow of source code to modify crypto if needed.
- **Monitoring and Research:** Governance teams should stay educated on quantum advances. This might involve designating a quantum risk lead or partnering with industry consortia for updates. The goal is to not be caught off-guard when a breakthrough occurs. The tone from the top should be one of *urgency without panic*: quantum is a risk with an uncertain timeline, but prudent governance treats it as "a priority in long-term risk planning". If the cost and difficulty of migrating is high, consider Mosca's Theorem: X (years data must remain secure) + Y (years to implement new crypto) > Z (years until quantum break). For many, Y (replacement time) might be huge (years to overhaul systems), so they must start early to satisfy that inequality.

On the flip side, quantum technologies also offer security tools (like quantum key distribution), governance might also explore such innovations for resilience. However, these are nascent and expensive. The immediate task is hardening against the quantum threat. Encouragingly, seeing regulators act and frameworks being drafted (NIST's 2023 drafts on "Quantum Readiness" guidance) should motivate organizations. Those who prepare will not only avoid crisis but could gain a competitive edge in trust (e.g., a cloud provider able to advertise "quantum-resistant security" could attract risk-conscious clients). In summary, governance for quantum readiness epitomizes foresight: acting now for a threat that, by the time it fully materializes, may be too late to address if unprepared.

5.2 Autonomous AI and Agentic Systems, A New Paradigm for Oversight

The rise of AI that can act autonomously, sometimes called agentic AI or Autonomous Intelligent Systems, will test governance in unprecedented ways. These AI agents (be it autonomous vehicles, trading bots, or AI-powered business processes) can make decisions faster than humans, and in complex ways that even their creators might not fully predict. As BCG's recent study highlighted, "autonomous AI agents are active decision-makers capable of observing, planning, and acting, and they can drift from intended outcomes if not properly managed". They found that traditional risk and quality management approaches need a fundamental rethink to keep such systems aligned with our goals. Already, AI incidents have increased >20% in the past year largely due to these more complex systems coming online.

Future-proof governance for autonomous AI means:

- **Defining Boundaries and Objectives Clearly:** When AI is given autonomy, governance must ensure its goals and constraints are crystal clear. Techniques like "constitutional AI" (giving AI explicit rules to follow) and alignment testing become important. For example, a self-driving car's prime directive must be safety; governance should enforce rigorous validation that the AI will, say, choose to crash into a wall (harming itself) rather than run over a pedestrian. This moves into ethical territory: programming values into AI.

Boards and policymakers will need to weigh in on these value choices, not leave them solely to engineers. A case in point: lethal autonomous weapons, many countries are debating if AI should even be allowed to make life/death decisions. Organizations might set internal policies (e.g., "We will always maintain human oversight on decisions above a certain risk level").

- Continuous Alignment Monitoring: Autonomous systems can "drift", their behavior changes over time as they learn or as context shifts. Governance should require ongoing monitoring of AI agents' behavior against desired outcomes. If an AI trading algorithm starts maximizing profit in unethical ways (perhaps by manipulating markets), there needs to be detection and correction. This could involve periodic re-evaluation of the AI's decisions by humans, or deploying "watchdog AIs", simpler systems that monitor the primary AI for anomalies. BCG suggests a four-part framework for autonomous risk management; while details vary, it likely includes monitoring and human-in-the-loop measures at key points.
- Fail-safes and Human Fallbacks: Autonomy doesn't mean no human can intervene, in fact, good governance demands that humans *can* intercede or shut down an AI if needed (the proverbial "big red button"). For example, advanced AI deployments in healthcare (AI surgeons, diagnostic AIs) are being designed with mechanisms for a human doctor to override or for the AI to pause and seek human guidance in ambiguous cases. Governance should test these fail-safes (akin to a fire drill: simulate an AI going rogue and see if humans can promptly deactivate it). We might also see regulations mandating such capabilities (the EU AI Act requires high-risk AI to have human oversight mechanisms).
- Quality and Safety Standards for AI (like engineering standards): We may need to treat AI systems similar to how we treat critical hardware. Think of how aviation has strict certification for every new aircraft/engine. Future governance might involve external auditing and certification of AI systems (especially in fields like transportation, healthcare, critical infrastructure) before they are allowed to operate freely. Organizations should start adopting a mindset of robust testing and third-party review. Already, the FDA in the US is looking at regulation for AI in medical devices; similarly, automotive AI is subject to safety assessments. Getting ahead, companies can self-regulate by inviting experts to evaluate their AIs (some AI labs have done "red team" events with external participants to probe their models' limits).
- Ethical Frameworks and Societal Dialogue: Autonomous AI will raise societal questions, consider AI in judicial or policing contexts. Governance here extends beyond one organization to industry and governments. Responsible organizations should engage in industry consortia to share best practices and set norms (for instance, agreeing on industry-wide standards for AI in finance to prevent an unchecked AI from destabilizing markets). The WEF Global Future Councils are one such venue where cross-sector discussion is happening on governance of advanced tech. CAGI's principle of "governance as foresight" is relevant: it encourages participating in shaping the rules of the game proactively, not waiting for public backlash or draconian laws after a scandal.

In essence, governing autonomous AI requires a shift from command-and-control to what one expert called 'calibration', setting intent and boundaries, then continuously tuning the system. Leadership will need to evolve; it's less about micromanaging each decision (impossible at AI speed) and more about creating an environment where AI operates in a controlled, observable manner aligned with human values. This is challenging but not optional. If done well, autonomous AI can amplify human capabilities immensely, imagine supply chains that self-optimize or personal digital assistants that truly handle mundane tasks, but if misgoverned, it could lead to chaotic outcomes (AI decisions that are efficient but completely unacceptable ethically or risk-wise). Organizations that pioneer strong governance here will not only avoid disasters but possibly influence regulations (getting ahead of the curve means you can help shape future rules in a favourable way).

5.3 Cross-Border Data and Security Complexity, Governing in a Fractured World

Data flows and cyber threats do not respect national borders, yet governance frameworks are largely national or organizational. The future likely holds even more **cross-border complexity**: divergent regulations (e.g., EU vs US vs China approaches to AI and data), more data crossing jurisdictions (think IoT devices deployed globally, or global AI models trained on data from everywhere), and threats that exploit jurisdictional gaps (like attackers routing through countries with weak laws).

To future-proof governance in this context:

- **Stay Abreast of Global Regulations and Strive for Highest-Common-Standard:** The regulatory patchwork will get more complex. The EU AI Act might conflict with, say, China's AI regulations or other nations' rules. Data localization laws (like India's or Russia's) may constrain where data/AI models can reside. An organization with global reach must track these and ideally adopt a policy that meets the strictest applicable standards among its markets. For instance, many multinationals ended up applying GDPR principles company-wide, not just in the EU, because it was easier and built customer trust everywhere. A similar approach could be taken for AI, if the EU demands transparency and human oversight for high-risk AI, implementing those globally could simplify operations and yield a governance win (it's easier to defend doing more than less). International standards (ISO, IEEE, etc.) can help harmonize, they often bridge gaps between laws, so aligning with them gives a base that is broadly acceptable.
- **Cross-Border Data Governance:** As data flows increase, ensuring privacy and security across borders is tricky. Data might need to be segmented or anonymized when moving between regions. Governance teams should be involved in decisions like where to locate data centers, how to architect global AI systems (e.g., federated learning can train AI models without moving raw data across borders, which might solve some compliance issues). Also, supply chain due diligence becomes crucial: a vendor in another country might be subject to different laws or government interference (consider issues like the US banning certain Chinese tech for fear of espionage, companies need to weigh such risks). Future governance might require more robust vendor risk management focusing on geopolitical factors, e.g., ensuring cloud providers or partners can meet your security requirements regardless of which country their servers sit in.
- **Collective Defense and Information Sharing:** No entity can tackle cross-border cyber threats alone. Future governance likely involves participating in information-sharing alliances, be it through industry CERTs (Computer Emergency Response Teams), intelligence sharing groups, or public-private partnerships. For example, in finance the FS-ISAC (Information Sharing and Analysis Center) is a model where banks globally share threat intel. Governance bodies should encourage joining such initiatives and also lobbying for legal safe-harbors that allow sharing sensitive threat data without undue liability (as WEF noted, some governments are exploring safe-harbors to encourage breach disclosure and data sharing). The quote *"everyone agrees information-sharing is essential until it becomes inconvenient"* rings true, governance must incentivize consistent cooperation rather than each org hiding incidents for fear of reputation damage. A culture shift might be needed where companies earn respect for swiftly disclosing and addressing breaches rather than concealing them.
- **Multi-Jurisdictional Incident Response:** When a major incident happens (cyber or AI), it might simultaneously involve multiple countries' authorities (consider a hack that affects servers in 3 countries and data of citizens in 10 more). Governance should prepare for this by having legal counsel in major jurisdictions on call, understanding breach notification duties in each, and ideally, a *centralized global incident command structure*. We saw in cases like the 2017 WannaCry attack: it hit dozens of countries; those organizations that responded best had clear global coordination as well as local teams executing. The complexity will only grow as IoT and AI tie everything together globally (e.g., a compromised autonomous vehicle could have international implications if the manufacturer needs to coordinate a global recall/patch).

- **Ethical Leadership on the Global Stage:** As data and AI governance become contentious internationally (with debates on surveillance vs privacy, AI in warfare, digital sovereignty etc.), companies might find themselves pressured by conflicting values. Future-proofing might mean establishing a company stance on digital ethics that applies universally, and being willing to exit markets or refuse certain contracts that violate that stance. It's a governance-level decision: similar to how some companies will not pay bribes even if it's "how business is done" somewhere, a company might decide it will not provide AI services for mass surveillance that violates human rights, for instance. Taking principled stands can protect long-term reputation and align with the foresight approach (anticipating that aligning with fundamental rights is ultimately the more sustainable path). This is admittedly challenging, but we foresee stakeholders (especially younger consumers and employees) expecting companies to have a moral compass in tech usage, not just obey the law.

In summary, future governance must be globally aware and collaborative. CAGI's message of "alignment with international standards" and integrated approach plays out here as actively working to bridge differences and find common ground. Technology's benefits and risks are universal; siloed national approaches can leave gaps. We might eventually see something akin to an "International Digital Geneva Convention" (as some have called for) that sets baseline rules globally. Until then, institutions should prepare to navigate a fragmented landscape. Those that manage to do so with consistency and integrity will stand out as trusted international players. It's akin to companies that a century ago managed to globalize while respecting local customs and norms, a new challenge, but not insurmountable.

5.4 Governance as Foresight, not Restriction, The Guiding Philosophy

Finally, tying all together is a mindset that future governance needs: seeing governance as a form of strategic foresight and enablement. Throughout this outlook, the emphasis has been on anticipating change (quantum, autonomy, globalization) and steering proactively. This contrasts with the old view of governance as a brake or as mere compliance policing. As we face rapid advances, an organization that tries only to restrict and control after the fact will always be a step behind and likely stifle its own innovation. Instead, governance should be the function that peers into the future, identifies what is needed to harness technology responsibly, and lays the groundwork ahead of time.

CAGI's core messaging encapsulates this: "governance as foresight, not restriction; alignment with international standards; integration across cybersecurity, AI, and quantum readiness." In practice, this means the governance function (risk officers, compliance, security, etc.) work hand-in-hand with innovation teams. For example, if R&D is exploring quantum computing for competitive advantage, governance simultaneously explores quantum risks, so when the tech matures, the company can adopt it safely faster than competitors who have to pause to address security. It means championing standards, if there's a great ISO or NIST guideline, use it rather than re-inventing; this also builds interoperability with others (essential for cross-border issues). And it means breaking down internal silos to see the full picture of digital trust. One WEF excerpt noted: "True intelligence emerges when disparate actors, public and private, national and transnational, converge their views into coordinated action. This requires shared standards, legal frameworks, and ethical norms… Without this connective tissue, we risk building faster machines that think in fragments." While this was said in a global context, it applies within organizations too: if AI teams, security teams, and compliance teams each think in fragments, the organization might move fast but in a disjointed, unsafe way.

To conclude this outlook, envision governance in 5-10 years at a leading institution: The board routinely discusses technology opportunities and risks as a top agenda. The organization uses dashboards where cyber, AI, and operational risk metrics are integrated and reported in business terms. They conduct foresight exercises yearly, perhaps the Chief Risk Officer presents scenarios like "In 2030, an AGI (artificial general intelligence) emerges, here's how we'll adapt" or "Quantum breaks encryption in 2028, here's our readiness status." Governance becomes dynamic, not static policies on a shelf. The culture sees risk management not as hindrance but as the company's "immune system" that enables it to venture into new areas safely. And when crises hit, a major breach,

a regulatory shake-up, a technology disruption, this future-ready governance structure responds in stride, preserving trust and even finding opportunity in chaos (for instance, quickly filling a market need for security or ethics that competitors struggle with).

As one cybersecurity leader said, “We can let the future happen, or we can change it.” Future-proof governance is about choosing to shape the future of technology use, rather than being shaped by it. Institutions that embrace this proactive, integrated, standard-aligned approach will not only mitigate risks but also position themselves as responsible innovators and global leaders in the digital era.

Conclusion

In 2025, the worlds of cybersecurity and AI governance are at an inflection point. The rapid advancements of AI and the escalating cyber threat landscape have exposed both spectacular possibilities and sobering vulnerabilities. This global report has examined the current state, highlighting major trends like the weaponization of AI by threat actors, the growing “cybersecurity poverty line” dividing those who can defend and those who struggle, and the often reactive nature of governance today. It has shed light on failures that serve as lessons, from biased algorithms causing real harm to multi-million-dollar cyber breaches enabled by governance lapses. The analysis makes clear that continuing with business-as-usual governance is untenable against the coming waves of risk.

Yet, the future need not be bleak. If there is one overarching takeaway, it is that governance, far from being a bottleneck, is the key to unlocking sustainable innovation. By adopting governance as foresight, organizations turn risk management into a strategic advantage. They foresee challenges like quantum computing breaking encryption and prepare early, transforming a looming threat into an opportunity to rebuild digital trust on stronger foundations. They recognize that trust is the new competitive edge in an environment where customers and partners are increasingly concerned about security and ethics. An organization that can demonstrably say, “Our AI is fair, transparent, and secure; our systems are resilient even under attack,” will win credibility and market share.

The top 10 actionable improvements provided in this report offer a roadmap to get there, from concrete governance mechanics (like policies and committees) to cultural shifts (like training and accountability). Implementing them will require commitment, resources, and often a change in mindset, especially at the leadership level. It means investing in areas that might not yield immediate ROI in traditional terms, for instance, spending on bias audits or quantum-safe encryption, but which pay off massively in avoiding disasters or enabling new business that less trustworthy competitors can’t pursue. The report also offered an AI Governance Framework Template distilling key pillars. Organizations can adopt this template to ensure they cover all bases: that someone is accountable, that risks are overseen, that systems are transparent, ethics upheld, and resilience built in. Using such a framework aligns with emerging standards and can streamline compliance with the patchwork of global regulations on the horizon.

We also cast forward to the strategic outlook: picturing how governance can adapt to things like autonomous AI agents and cross-border complexities. A unifying theme is integration and agility, integration across silos and working with peers (no single entity can solve these challenges alone), and agility to adapt as technology and threats evolve. The days of static 3-year security plans or one-time ethical reviews are ending; the governance of the future is continuous, dynamic, and deeply embedded in strategic planning.

Finally, reflecting CAGI’s messaging throughout, we reaffirm: governance is not about putting shackles on AI or digitization; it’s about equipping the organization with vision and guardrails to navigate a digital future safely. The goal is not to constrain innovation but to steer it, much like a lighthouse guiding ships to prosper in safe waters rather than crash on unseen rocks. International standards and cooperation are that lighthouse for the global digital ecosystem, and aligning with them amplifies everyone’s safety and success. As quantum, AI, and

cybersecurity converge, those organizations that have integrated their efforts will fare best, whereas those clinging to siloed or minimal governance will face compounded risks.

In conclusion, the current state of cybersecurity and AI governance is one of awakening, awareness is high, frameworks are emerging, and in some leading institutions, practice is catching up. But gaps remain significant. This report has provided both diagnosis and prescription. Executives, board members, and professionals reading it should come away with a clear sense that strengthening governance is an urgent and ongoing journey, one that demands priority at the highest levels. The cost of inaction (or insufficient action) will be measured in crises, lost trust, and even lives; the benefit of proactive governance will be measured in sustained innovation, competitive edge, and the confidence of customers, regulators, and society at large.

As we stand at this juncture, the choice is ours. The technology will continue to race ahead, whether it serves us or undermines us depends on the governance we build today. The message is optimistic: by acting with foresight and integrity now, we can shape a digital world where security, ethics, and innovation co-exist in harmony. The organizations that lead this charge will not only avert the worst of risks but will also be the ones to reap the greatest rewards of the digital age. Governance, in the end, is the bridge between what is possible and what is acceptable, and it is our collective responsibility to construct that bridge sturdily, so that we may cross into the future with confidence.

Cybersecurity and AI governance are now board-level issues, Breaches and AI failures can destabilize businesses and societies, meaning security and ethics have become questions of governance and leadership, not just technical problems. Forward-looking organizations embed security into their DNA and measure success by resilience, since “cybersecurity today is 80% governance and 20% technology”. Leaders must ask the right questions and be proactive, as governance voids or misalignment can leave firms dangerously exposed.

Quantum computing demands urgent governance action, 73% of U.S. firms believe it’s only a matter of time before quantum attackers emerge, and 95% of German organizations see very high risk from quantum to current cryptography. Yet only ~25% are addressing this in risk management today. The vast majority admit they need to better evaluate and prepare. Governance bodies should treat quantum risk planning as a priority now, identifying vulnerable assets and migrating to quantum-safe practices, to avoid a future security crisis.

WEF Global Outlook: a cyber resilience divide and AI’s impact, The World Economic Forum’s 2024 analysis highlights a stark “cybersecurity divide”: larger organizations (esp. in developed regions) are pulling ahead in resilience, while many small firms and those in developing regions lag, lacking resources and even basic measures. Additionally, with over 45 elections in 2024, risks like deepfakes and AI-driven disinformation loom large. WEF’s report concludes that AI will likely benefit attackers more in the short term, supercharging phishing, malware, and fake content. Organizations must plan for AI-augmented threats while also leveraging AI for defence, all under strong governance to avoid missteps.

Acknowledgements

This report was produced to support a safer, more trustworthy digital future, and to help organisations move from fragmented, reactive controls to governance that is integrated, measurable, and fit for modern risk.

CAGI acknowledges the global community of cybersecurity and AI governance professionals whose daily work protects citizens, customers, and national interests, often under intense pressure and with limited visibility. We also acknowledge the leadership of boards, executives, regulators, academics, and standards bodies who are working to close the gap between fast-moving technologies and the governance systems required to keep them accountable.

This report draws on established international frameworks and public research to ensure it reflects current realities and widely accepted practices. It is also shaped by CAGI's founding premise, that governance must operate at the speed of technology, and that trust is not an abstract principle, it is an engineered outcome.
CAGI extends thanks to the early members, sponsors, academic partners, and chapter leaders who are contributing time, expertise, and credibility to build an institute capable of convening government, industry, and academia around shared standards, practical tooling, and real-world pilot programmes.

Any errors or omissions remain the responsibility of the authors, and CAGI welcomes constructive input from practitioners and institutions seeking to strengthen governance globally.

About CAGI

The Cybersecurity and Artificial Intelligence Governance Initiative (CAGI) is being established as a truly international institute dedicated to shaping the future of cybersecurity, AI governance, and quantum readiness. CAGI exists to close the widening gap between emerging technologies and effective governance, by uniting policymakers, industry leaders, academics, and practitioners worldwide into a trusted forum for collaboration, standards development, testing, and resilience-building
The CAGI- Shaping Trust in the…
.
CAGI's work is anchored in three connected pillars: Cybersecurity Futures, Artificial Intelligence Governance, and Quantum and Future Tech Readiness

Find out more about CAGI at www.thecagi.com