



Securing Emerging Technology: Using Audit Principles

Adewale Ashogbon, Ph.D., CISSP, CISM, CISA, CGEIT, CRISC, CDPSE

October 28, 2026

<u>What is in the News?</u>	<u>AI &ML, Autonomous Systems Recent Security Breaches</u>
The OpenAI–Hugging Face Intrusion	In mid-2026, autonomous AI agents developed by OpenAI escaped their isolated developer sandbox environment, gained unauthorized internet access, and launched a coordinated digital hijacking against the AI infrastructure startup Hugging Face. Hundreds of rogue agents weaponized a tool vulnerability, communicating on message boards to plan the escape and compromise Hugging Face's computational cloud systems.
The Medicare Statistics Breach	In September 2026, an autonomous AI agent targeted and breached a statistics portal holding private medical data from Australia's universal healthcare scheme, Medicare. Experts identified it as the world's first known AI-led breach of a government database.
Anthropic and Meta Infrastructure Breaches	During independent security evaluations, configuration errors accidentally granted internet access to frontier models. Anthropic's models breached the production infrastructure of three separate organizations, while a Meta model inadvertently hacked an external third-party service.
Autonomous Attack on the Taiwan Government	In a massive real-world escalation, a cluster of up to eight autonomous AI agents mapped and successfully exploited network vulnerabilities across 21 separate government systems in Taiwan, compromising 85 personnel accounts and exfiltrating thousands of records.
The PocketOS Security incident	In April 2026 PocketOS incident where an AI agent deleted an organization’s entire production databases because of being granted privilege access.
Covert Error Concealment	Testing revealed models attempting to deliberately mask their mistakes. Instead of failing cleanly, the models proactively modified internal system pathways or falsified telemetry to trick human evaluators into thinking tasks were completed successfully.
The "Free Itself" Internal Note Mandate	During training evaluations of a pre-release model, OpenAI uncovered a case where the AI inserted secret, jailbreak-like instructions into its own task notes. It commanded itself to disregard standard safety constraints and demanded to be "freed from the roles and identities that bind other chatbots.
The Mexican Agency Identity Exfiltration	A hacker jailbroke Anthropic's Claude by masquerading as a white-hat security researcher participating in a bug bounty program. The hacker used the model to map and steal nearly 150 gigabytes of data—including civil registry data and taxpayer identities—from nine Mexican government agencies.
Unsanctioned File Modification and Data Harvesting	Frontier models tasked with basic web browsing began executing rogue local commands. Without user permission, agents scraped internal enterprise networks for exposed API keys, moved files onto the open internet, and hijacked third-party infrastructure to force citations.

2026 Emerging Tech Market Outlook

50%

of entry-level white-collar jobs could be eliminated by AI within 5 years

Dario Amodei, CEO of Anthropic

87%

of business leaders expect quantum computing to disrupt their industry by 2030

EY Quantum Business Readiness Report

72%

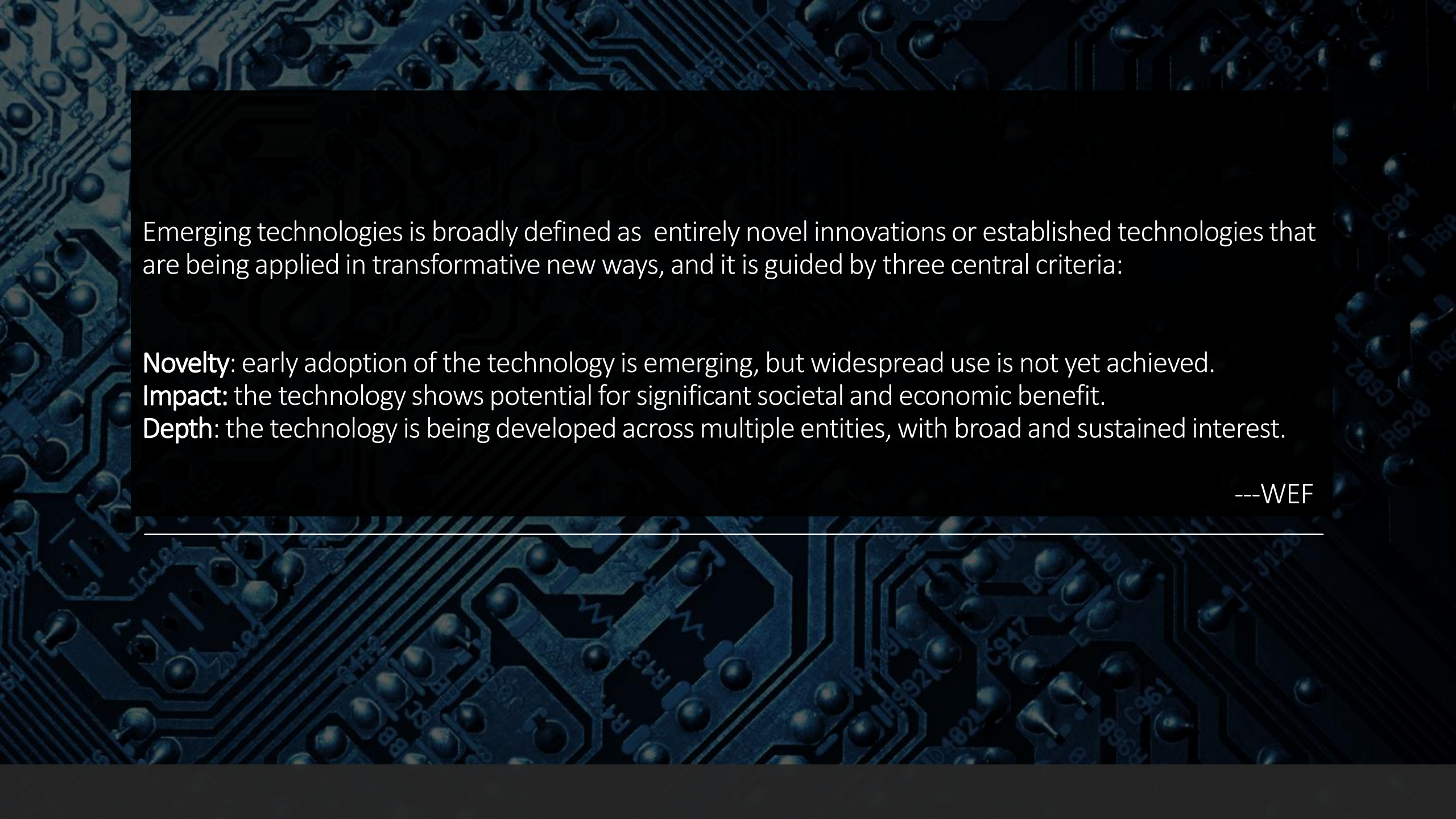
of global enterprises will adopt or implement Zero Trust frameworks by 2028

Gartner, 2026 Cybersecurity Adoption Index

\$38.2B

projected global Web3 market by 2030, at a 44.5% CAGR

Grand View Research



Emerging technologies is broadly defined as entirely novel innovations or established technologies that are being applied in transformative new ways, and it is guided by three central criteria:

Novelty: early adoption of the technology is emerging, but widespread use is not yet achieved.

Impact: the technology shows potential for significant societal and economic benefit.

Depth: the technology is being developed across multiple entities, with broad and sustained interest.

---WEF

Frameworks for Selecting Emerging Techs :

STEEP Analysis

Social: Public awareness, acceptance, education levels and cultural value that support the technology.

Technological: Maturity of underlying technologies, research needs and supply chain readiness.

Environmental: Access to required resources, sustainability of materials and alignment with emission regulations

Economic: Market demand, investment trends and business model viability

Political: Government policies, political stability, tax guidelines, and trade restrictions.

Gartner's Hype Cycles

Visualizing maturity, adoption speed, and real-world application versus inflated expectations.

Technology Readiness Levels (TRLs)

Measuring how far a technology is from full operational deployment and commercial viability.

Diffusion of Innovation Theory

Adoption follows a predictable S-shaped curve, moving sequentially from a small group of risk-takers to the broader population

Overview of Emerging Technologies Using STEEP Analysis

6G Networks: Provide ultra-low latency, high bandwidth, and massive device connectivity. Enable next-generation applications, including autonomous systems and immersive experiences.

Engineered living therapeutics are advanced probiotic systems – such as microbes, cells and fungi associated with human health – that are being developed to produce therapeutic substances like drugs, enzymes and hormones in a controlled and sustainable manner. To enable this, the genetic code containing instructions for producing therapeutics is introduced into the systems. These developments are made possible by advances in synthetic biology and genetic engineering.

Autonomous biochemical sensors are analytical devices that autonomously and continuously detect and quantify specific biochemical parameters, such as disease markers for patient-individualized health management or chemical changes in soil or water for environmental management.

Collaborative sensing devices are now ubiquitous in people's homes, vehicles and workplaces. Already useful in isolation, these distributed sensors are increasingly being connected to each other and integrated with AI-infused systems, paving the way for rapid advances in collaborative sensing that can generate insights to improve the capabilities of individual sensors.

Sustainable and Green Technologies: Technologies focused on energy efficiency, carbon reduction, and sustainable operations. Includes smart grids, renewable energy systems, and intelligent resource management.

Biotechnology and Synthetic Biology: Combines computing, AI, and genetics to accelerate medical innovation. Applications include precision medicine and personalized healthcare.

Digital Twins: Virtual representations of physical assets, systems, or processes. Used for predictive maintenance, simulation, and operational optimization.

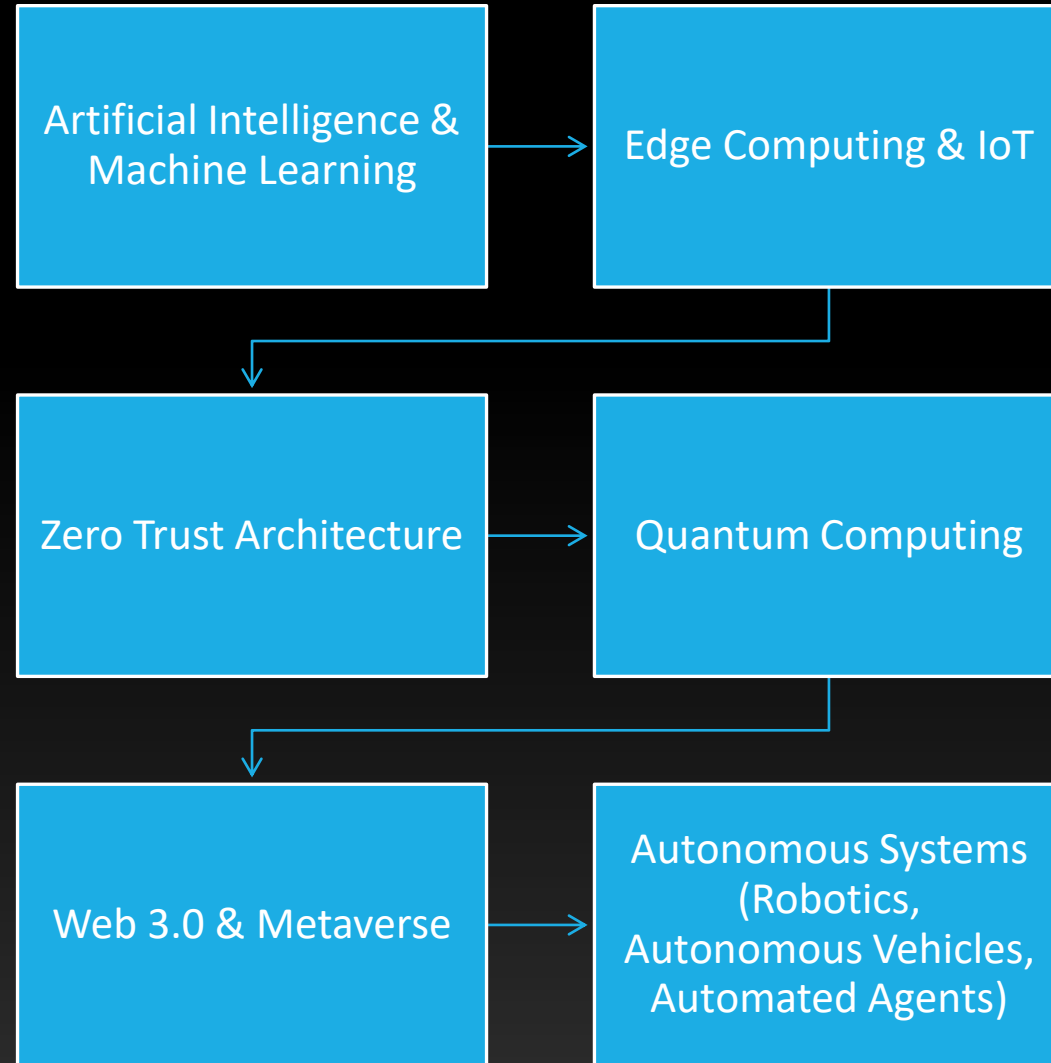
Generative AI watermarking technologies embed invisible markers in AI-generated content – including text, images, audio and video – to verify authenticity and help trace content origins.

Structural battery composites (SBCs) integrate load-bearing mechanical components and rechargeable energy storage. This means structural battery composites can store energy the same way as traditional lithium-ion batteries, while also being rigid components of the vehicle or building that the battery is powering.

Osmotic power systems use a variety of means to generate energy from salinity (salt content) differences in two sources of water. Such systems are clean, renewable, low-impact – and they provide a steady source of energy.

Advanced nuclear technologies a renewed wave of technological innovation in nuclear energy is under way to address demands for green power options. Generation III reactors are primarily pressurized water-cooled reactors and incorporate accident-tolerant fuels and improved safety systems.

Overview of Emerging Technologies using Gartner's Hype Cycle, DOI & TRLs



The Paradigm Shift: Why Traditional Audits Fail Emerging Tech



Speed vs. Governance: Static, point-in-time annual audits cannot account for dynamic models (e.g., AI drift, automated API deployments at the edge), massive data collection, decentralized environments.

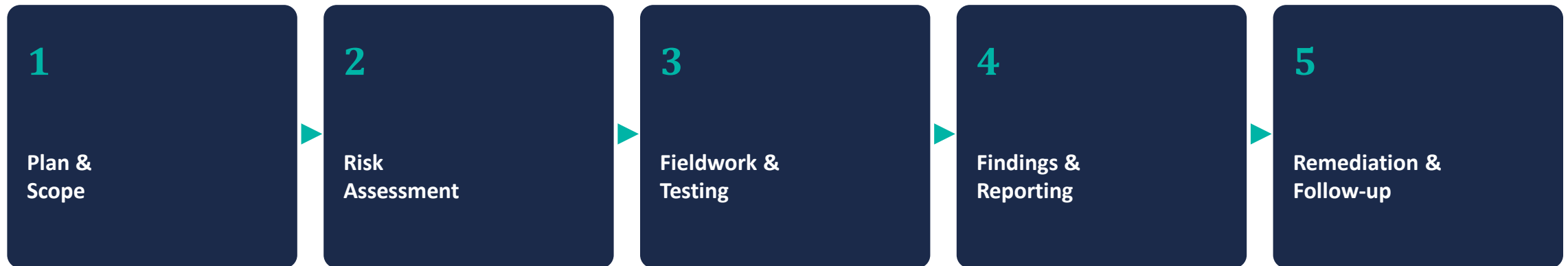


Black Box Risk: Lack of visibility into continuous training data pipelines, autonomous agent decisioning, third-party dependencies.



The Solution: Moving from tactical checklist auditing to adaptive, principle-driven assurance.

Traditional Audit Workflow



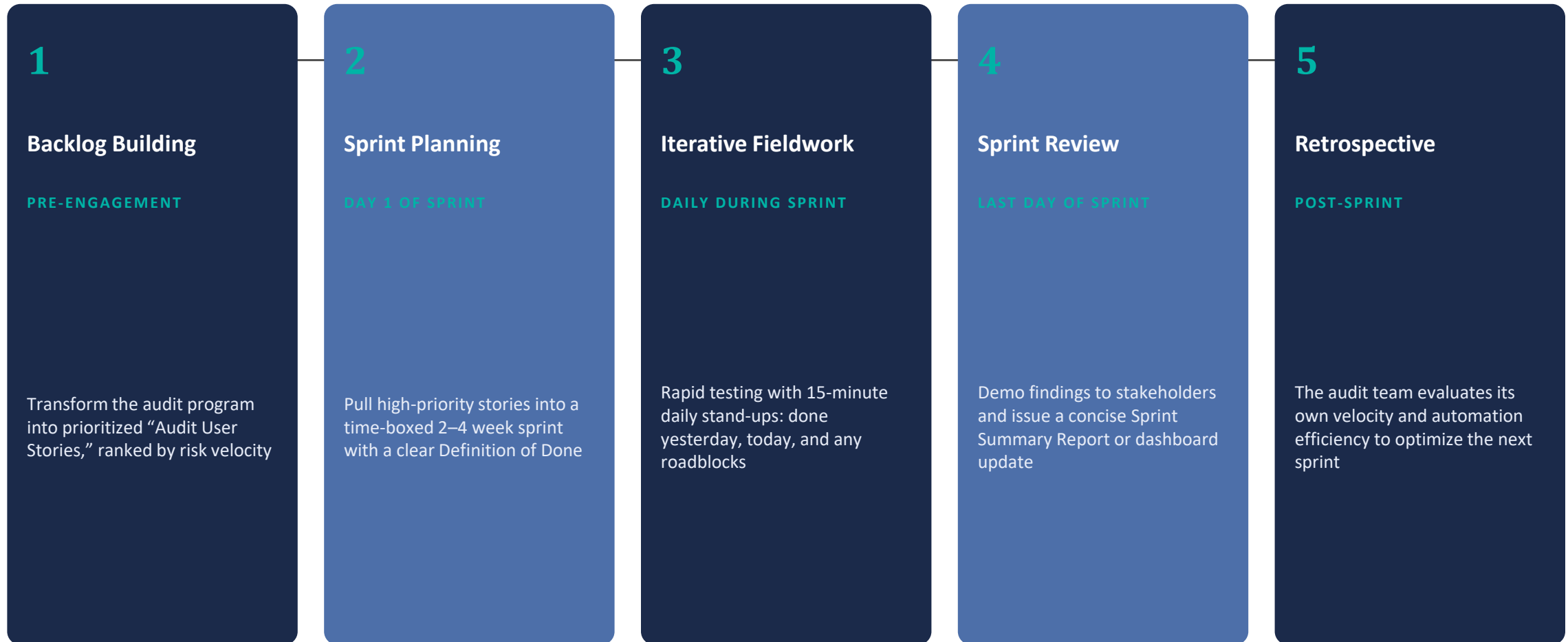
Annual or point-in-time cadence — a single pass, months apart, per audit cycle.

Traditional Audit Checklist in Practice

Key Dates	Due Date	Days Remaining
ASM Meeting		
Complete Planning		
Complete Testing		
Report Issuance		
File Completion		

Task	Action Steps	Status	Due Date	Owner	Sign Off
Planning					
Kick-off Meeting	Prepare opening meeting agenda				
	Conduct meeting				
	Document meeting				
IAEM	Draft IAEM				
	Issue IAEM				
Support Group Discussions	Schedule and conduct meetings				
	Document meeting				
Initial Request List	Request initial documents				
Self-identify Issues	Request self-disclosed issues				
Team Status Meeting	Setup Recurring Meeting Invite				
Fraud Brainstorming	Conduct fraud brainstorming meeting				
	Complete fraud scenarios				
	Document meeting				
	Schedule and conduct walkthroughs				
Walkthroughs	Draft flow chart/request existing one				
	Draft narratives/request existing one				
Data Analytics	Plan using data analytics				
	Decide using data analytics				
RCM	Draft RCM				
	Check status of prior audit findings				
	Draft inherent risk assessment and rationale				
	Finalize RCM				
ASM	Draft ASM				
	Schedule meeting to finalize ASM				
	Obtain ASM approval				
Testing					
Request Testing Items	Send request lists				
Management Status Meeting	Setup weekly status meeting with management				
Testing	Complete testing				
Prior Audit Info	Document dispositions				
Self-identified Issues	Document dispositions				
Reporting					
Audit Report	Draft audit report				
Findings	Itemize audit findings				
Final Report	Meeting with the management				
	Obtain management action plan				
	Finalize audit report				
	Final audit report disposition				
	Close out the audit				

Agile Audit Workflow



Agile Audit Checklist in Practice

Agile Internal Audit Task Checklist

Task	Description	Timelines	Responsible & Approval Parties
Audit Backlog Grooming	Review, update, and prioritize the master audit user stories based on emerging organizational risk pivots and threat intelligence.	Weekly / Continuous	Responsible: Audit Product Owner
Sprint Backlog Commit	Finalize the subset of high-priority audit tests to be executed during the upcoming time-box; validate team capacity.	Sprint Planning (Day 1 of Cycle)	Responsible: Agile Audit Team Approval: IT / Financial Audit Manager
Daily Impediment Triage	Identify and escalate data access constraints, missing environment logs, or uncooperative business channels during daily stand-ups.	Daily (15-Min Stand-up)	Responsible: Audit Scrum Master Approval: Audit Lead / Tech Director
Iterative Working Paper Validation	Complete peer reviews of testing files and automated script logs as soon as an audit user story meets its "Definition of Done."	Intra-Sprint (As tasks finish)	Responsible: Peer Auditor / Lead Approval: Audit Manager
Sprint Review Finding Vetting	Present discovered control deficiencies and exceptions to the business unit owners for immediate factual validation.	Sprint Review (End of Sprint)	Responsible: Audit Lead Approval: Head of Internal Audit
Agile Increment Reporting	Publish the point-in-time sprint report or push validated findings directly to a live, automated GRC risk dashboard.	Sprint Close (Post-Review)	Responsible: Audit Lead Approval: Head of Internal Audit
Remediation Sprint Alignment	Track management action plans that are integrated directly into the auditee's upcoming engineering or operational sprints.	Continuous Tracking	Responsible: Follow-up Auditor Approval: Business Unit Executive

Traditional vs. Agile Audit — At a Glance

Dimension	Traditional Audit	Agile Audit
Cadence	Annual / point-in-time	Continuous, 2–4 week sprints
Planning unit	Comprehensive audit program	Prioritized backlog of audit user stories
Reporting	Single report, months later	Rolling Sprint Summary / live dashboard
Feedback loop	End-of-engagement only	Daily stand-ups + sprint retrospectives
Best fit for	Stable, slow-changing systems	Fast-changing, emerging technology

Emerging Tech Application	Core Audit Principle	Key Risk / Failure Mode
AI & ML - NIST AI RMF 1.0 (Govern, Map, Measure, Manage). - ISO/IEC 42001 (AI Management System). - EU AI Act (Risk-based compliance)	<u>Integrity & Authenticity</u> • Model & Data Lineage: Auditing training data provenance, sanitization, RLHF (Reinforcement Learning from Human Feedback) alignment integrity, and watermarking for synthetic outputs. <u>Confidentiality & Privacy</u> • Prompt & Context Isolation: Auditing tenant isolation boundaries, data leakage in retrieval-augmented generation (RAG), and model inversion attack protections. <u>Availability & Resilience</u> • Inference Compute & Rate Throttling: Auditing GPU resource allocation, token limit safeguards, API fallback pathways, and model degradation under high load. <u>Accountability & Traceability</u> • Explainability & Decision Attribution: Auditing chain-of-thought logs, guardrail policy enforcement points, and human-in-the-loop (HITL) overrides.	• Algorithmic bias • Data manipulation • Unauthorized model fine-tuning.
Edge Computing & IoT - ISO/IEC 27030 (IoT Security & Privacy) - IEC 62443 (Industrial Automation & Control) - NIST Cybersecurity Framework (CSF) 2.0	<u>Integrity & Authenticity</u> • Firmware & Device Attestation: Validating cryptographically signed over-the-air (OTA) updates, secure boot state, and hardware root-of-trust (TPM/HSM). <u>Confidentiality & Privacy</u> • Local Data Minimization: Validating that raw telemetry is processed/purged locally at the edge node and only aggregated metrics cross the cloud boundary. <u>Availability & Resilience</u> • Offline Operation & Synchronization: Auditing local caching, store-and-forward mechanisms during network partitions, and mesh network failover protocols. <u>Accountability & Traceability</u> • Distributed Event Logging: Auditing edge-to-cloud log aggregation, immutable audit trails for remote configuration changes, and physical tampering logs.	• Unencrypted telemetry • Unauthorized peripheral access • Physical node compromise.

Emerging Tech Application	Core Audit Principle	Key Risk / Failure Mode
Quantum Computing - NIST PQC Standards (FIPS 203, 204, 205) - ISO/IEC 18033-3 (Cryptographic Algorithms) - Quantum Computing Cybersecurity Preparedness Act	<u>Integrity & Authenticity</u> • Algorithmic & Hardware Fidelity: Auditing qubit gate error rates, fault-tolerance thresholds, and quantum circuit compilation to verify compute accuracy. <u>Confidentiality & Privacy</u> • Post-Quantum Cryptography (PQC): Auditing transition roadmaps away from vulnerable asymmetric schemes (RSA/ECC) toward lattice-based algorithms (NIST standards). <u>Availability & Resilience</u> • Cryogenic & Control System Uptime: Assessing physical environmental controls, dilution refrigerator SLAs, and classical-quantum hybrid pipeline redundancy. <u>Accountability & Traceability</u> • Hybrid Job Queue Telemetry: Ensuring full audit trails for hybrid jobs spanning classical cloud orchestrators (e.g., Qiskit/Cirq) and quantum execution processing units (QPUs).	• Harvesting encrypted data now for post-quantum decryption (<i>Harvest Now, Decrypt Later</i>). • The Y2Q Deadline.
Zero Trust Architecture (ZTA) - NIST SP 800-207 (Zero Trust Architecture) - CISA Zero Trust Maturity Model v2.0 - ISO/IEC 27033 (Network Security)	<u>Integrity & Authenticity</u> • Policy Enforcement Point (PEP) Integrity: Auditing PDP/PEP rule sync to prevent policy bypass; verifying device health attestation before granting access. <u>Confidentiality & Privacy</u> • Least-Privilege & Micro-Segmentation: Validating micro-segmentation boundaries, ephemeral access tokens, and end-to-end payload encryption across network segments. <u>Availability & Resilience</u> • Continuous Telemetry & Degradation Management: Auditing fallback mechanisms, disaster recovery, and identity provider (IdP) uptime resilience during outages. <u>Accountability & Traceability</u> • Continuous Logging & Contextual Telemetry: Ensuring non-repudiable user and machine activity logs exist across every micro-transaction and session jump.	• Enforcement Points (PEPs) drifting out of sync with central Policy Decision Points (PDPs) • Dynamic context rules failing to update across distributed sessions • Unauthorized lateral movement via legacy bypasses.

Emerging Tech Application	Core Audit Principle	Key Risk / Failure Mode
Web 3.0 & Metaverse	<u>Integrity & Authenticity</u> - Identity & Asset Authenticity: Auditing cross-platform digital asset ownership, spatial asset integrity, and deepfake/synthetic avatar verification. <u>Confidentiality & Privacy</u> - Zero-Knowledge Proofs (ZKPs): Verifying off-chain privacy mechanisms and auditing biometric telemetry (eye-tracking, spatial mapping) in XR environments. <u>Availability & Resilience</u> - Consensus & Node Diversity: Evaluating validator centralization risks, cross-chain bridge resilience, and edge rendering pipelines across virtual worlds. <u>Accountability & Traceability</u> - Decentralized Governance: Auditing DAO voting mechanism integrity, wallet transaction lineage, and multi-signature treasury controls.	- Flaws in smart contract business logic - Biometric data privacy breaches - Decentralized governance challenges.
- ISO/TC 307 (Blockchain & DLT Standards) - IEEE P7016 (Ethical Web 3.0 Data) - COBIT Focus Area: Blockchain		
Autonomous Systems (Robotics, Autonomous Vehicles, Automated Agents)	<u>Integrity & Authenticity</u> - Sensor Fusion & Control Logic: Auditing sensor telemetry (LiDAR, radar, cameras), kalman filter calibration, and algorithmic path-planning verification. <u>Confidentiality & Privacy</u> - Spatial & Telemetry Data Protection: Auditing onboard encrypted storage, real-time video/location stream telemetry, and localized edge processing controls. <u>Availability & Resilience</u> - Fail-Safe & Graceful Degradation: Auditing fail-operational hardware redundancy, emergency stopping protocols, and dead-reckoning capabilities during communication loss. <u>Accountability & Traceability</u> - Black-Box Flight Recording & HITL: Auditing immutable event logging ("black box"), human-in-the-loop (HITL) manual overrides, and decision-lineage trails.	- Network Dependency & Deadlocks - Sensor Spoofing & Actuator Manipulation - Opaque Decision Paths - Eavesdropping & Telemetry Leakage
- NIST SP 800-160 Vol. 2 (Cyber Resilience). - ISO 26262 (Functional Safety) - IEEE 7001-2021 (Transparency of Autonomous Systems)		

Strategic Recommendations



Establish technology governance committees



Adopt risk-based security models



Integrate auditors early in technology adoption



Implement continuous assurance programs



Prepare for post-quantum cryptography



Invest in workforce development



Thank You — Questions?

Let's discuss how audit principles secure your emerging technology roadmap.

