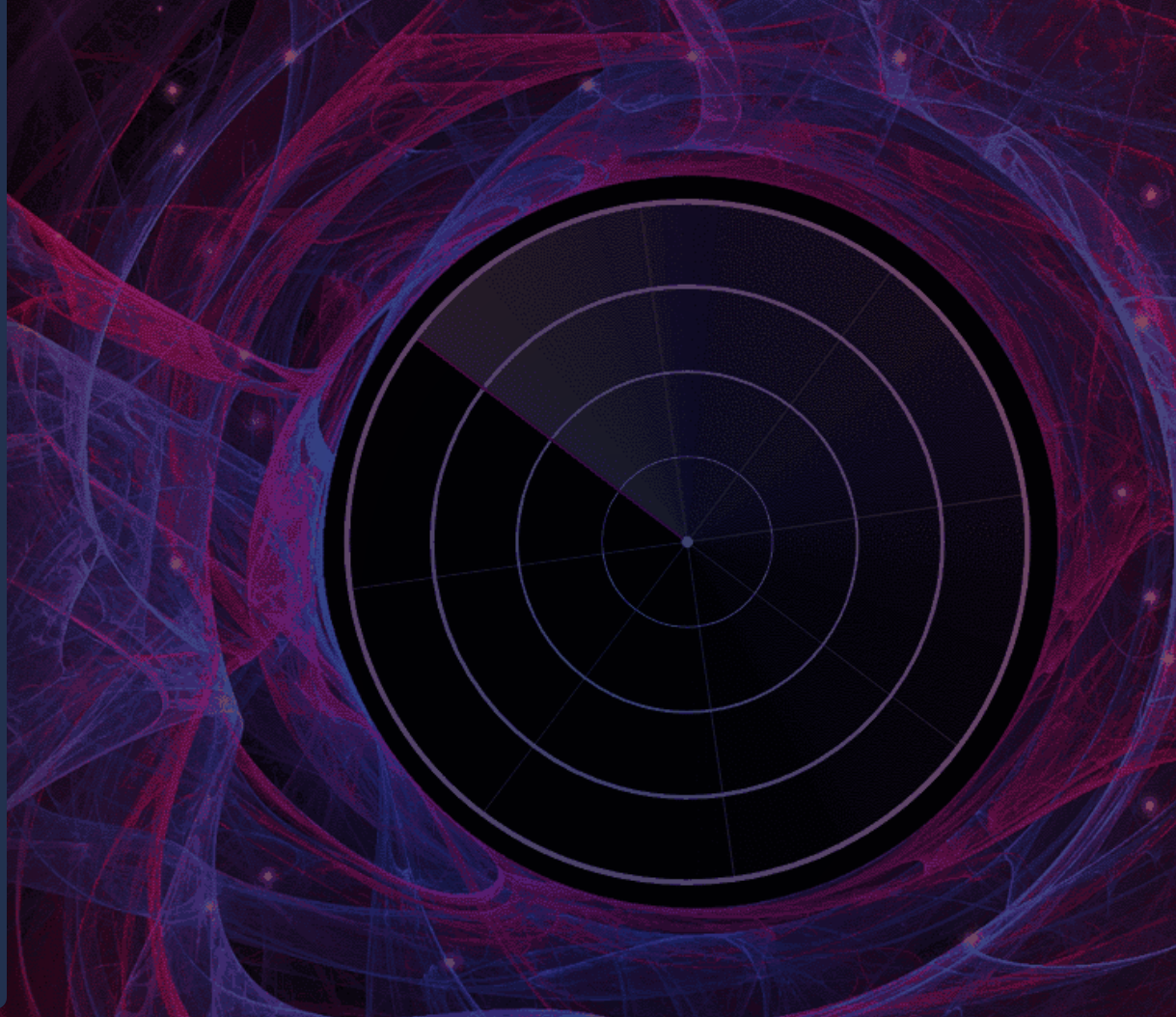


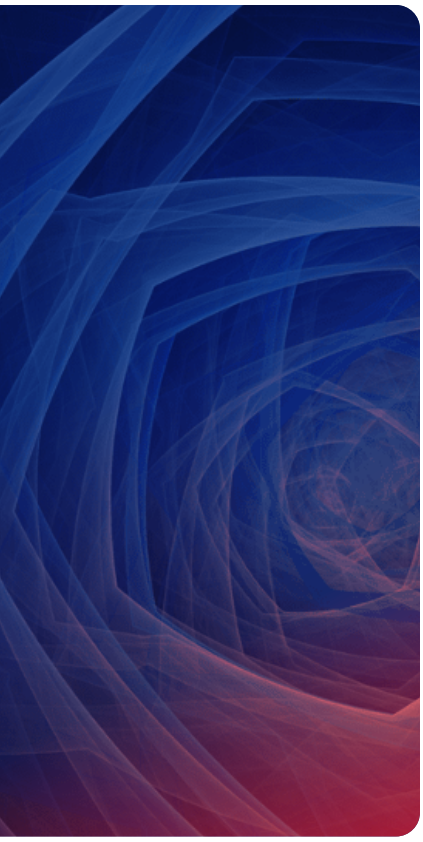
EBOOK

SAFETY FIRST:
PREPARING FOR THE AGE OF AI

A PRACTICAL GUIDE TO NAVIGATING AI RISK, READINESS, AND USE CASE ADOPTION

AVERTIUM®

CONTENTS



THE TWO FACES OF AI: INNOVATION FOR DEFENSE – AND EXPLOITATION

Artificial intelligence has many faces. Some are dazzling, promising unprecedented speed, insight, and innovation. Others are far more unsettling. The truth is, AI isn't inherently good or bad. It is, however, a force multiplier. And in cybersecurity, such an accelerator cuts both ways.

On the one hand, organizations are using machine learning and AI-driven systems to solve problems that were previously out of reach. They're predicting outages before customers ever feel the impact. They're stitching together fragmented telemetry into unified insights in seconds. And they're offloading tedious, repetitive tasks to intelligent agents that free their most valuable talent to focus on strategy.

Some companies are already saving thousands of dollars and reclaiming countless hours by allowing AI agents to autonomously handle password resets, access requests, and basic troubleshooting without a single human stepping in. IDC predicts that by late 2026, 65% of organizations will rely on AI-driven assistants, advisors, and agents to deliver immediate business and employee value, transforming the way IT operates.[1] These agents won't just respond – they'll *learn*, *reason*, *and act*, keeping systems running with minimal human intervention. It's a fundamental shift from reactive operations to a new era of intelligent, autonomous IT.

But as with any force multiplier, there's a flip side.

AI IS A DOUBLE-EDGED SWORD, AND BOTH EDGES ARE SHARP

For every groundbreaking efficiency AI introduces, equally powerful risks are emerging from the shadows. Threat actors are accelerating just as quickly – if not faster – to weaponize the same technologies, increasing attack volume and complexity, evading defenses, and exploiting human trust at unprecedented speed. Threat actors have wasted no time weaponizing AI to:

- Craft highly convincing phishing emails tailored to specific targets.
- Generate deepfake audio and video to impersonate executives.
- Mutate malware in near real time to escape detection.
- Map network topologies and probe systems autonomously.

What's more, as AI grows more autonomous, the stakes grow too. **Just like people, agents can be deceived, confused, or steered off course**. And as organizations start to accumulate more agents – IDC predicts 1.3 billion AI agents will be in circulation by 2028[2] – so does the potential for shadow agents, inconsistent governance, and widening blind spots across the attack surface.

This poses some serious questions for every IT and security leader: How do you choose the right AI use cases, set the right guardrails, and innovate fast enough to stay competitive without widening your attack surface and introducing more risk? Where should AI act on your behalf – and where must humans remain firmly in the loop? In the chapters ahead, we'll break down how to navigate this new territory and offer a perspective for responsibly adopting and transforming operations across your organization with AI.

HOLDING AI ACCOUNTABLE WITH HUMANS IN THE LOOP

Organizations that are successfully leveraging AI as a competitive differentiator aren't doing so by replacing people. They're redefining how people and machines work together. They're building hybrid teams where intelligent agents and human experts operate side by side. This requires a thoughtful reimagining of how responsibilities are divided – deciding which tasks can be reliably entrusted to AI and which still demand the discernment, judgment, and context only human experts can provide. The trick is figuring out which tasks you should delegate to AI and when humans need to intercede to mitigate potential risks.

WHERE AI EXCELS

AI excels at scale. It can process large volumes of data in seconds – far beyond what any human team could manage. It also performs exceptionally well when the task is repetitive, data-heavy, and time-sensitive. In practice, this means AI is excellent at duties like summarizing large volumes of information; identifying trends, anomalies, or relationships across different data sources; classifying and categorizing data; assisting with high-volume triage and prioritization; compiling insights in reports; and continuously monitoring systems in real time.

WHERE HUMAN EXPERTISE IS CRITICALLY NEEDED

When decisions carry meaningful financial, operational, or reputational consequences, it's best to let the human professionals take the lead. AI struggles when context spans multiple domains – technical, business, regulatory, and human. It's not adept at intuitively weighing organizational risk against operational urgency. And it does not naturally navigate scenarios where the "right" answer lives in a gray area rather than a binary choice. Put it in this perspective: If taking down a server costs a million dollars an hour, you shouldn't leave that decision to an autonomous AI agent.

TREATING AI AGENTS AS GOVERNED DIGITAL WORKERS

Even if you've meticulously defined the roles of your AI and human workers, you still need to establish additional guardrails. This is where another mindset shift comes in: AI agents should be treated as digital workers with defined responsibilities, access levels, and behavioral expectations. Put simply, **if agents are going to share the workload, they must also share the accountability**. That means applying the same level of oversight we extend to human users – governing agents through identity, access control, auditability, boundaries, and a clearly articulated purpose.

If we're going to accept the reality of hybrid human and agent teams, we must also accept the responsibility that comes with them. Agents need:

- **An owner** – someone accountable for what the agent does.
- **Clear boundaries** – a defined scope for where the agent can and cannot operate.
- **Auditable activity** – visibility into the actions an agent takes and why.
- **Appropriate permissions** – access levels aligned to its purpose, not more.
- **Behavioral guardrails** – policies that shape how the agent should behave in normal and exceptional situations.

These are the same fundamentals we rely on to manage human users since the risks are similar. Agents can be misconfigured, manipulated, over-permissioned, or misunderstood if we don't govern them with intention. So, the next question becomes: How do you prepare your organization for AI in a way that empowers the people at the heart of your business to use it responsibly – and unlock meaningful productivity gains along the way?

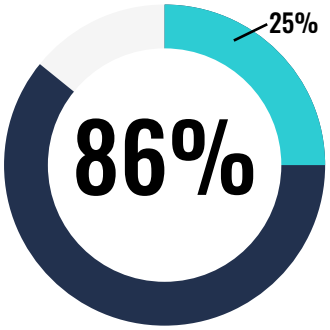
FOUNDATIONS OF AI READINESS: A PRACTICAL GUIDE

Meaningful AI transformation isn't achieved simply by deploying new copilots, agents, or use cases. The real transformation begins earlier – with readiness. Before an organization introduces AI into its environment, it must ensure the foundation beneath it is stable, governed, and prepared to support a new way of working.

At Avertium, we think about AI readiness through three interconnected pillars: AI governance, technical enablement, and data-centric controls. Together, these pillars create the conditions where AI can be adopted safely, responsibly, and with confidence – without undermining trust, security, or business operations.

1 AI GOVERNANCE: SETTING DIRECTION, ACCOUNTABILITY, AND BOUNDARIES

AI governance is where readiness begins. It brings structure to how AI is introduced – clarifying accountability, aligning on desired outcomes, and establishing a framework that allows organizations to scale AI use cases without introducing risk. Today, many organizations are innovating faster than they are putting guardrails in place, creating a widening gap between experimentation and accountability. Everyone wants to move fast with AI, but a word of caution: don't let speed compromise responsible implementation.



86% of organizations are aware of AI regulations, but only 25% have a fully implemented AI governance program.[3]

What does strong governance look like? To prepare for AI, organizations must lay the groundwork for a successful end state:

- Gain leadership buy-in to drive specific outcomes.
- Define desired outcomes and develop associated use cases.
- Understand current state and what achieving your desired outcomes requires.
- Establish policies and procedures to define acceptable use of emerging technologies to enable organizational goals.

To make governance practical, many organizations lean on established frameworks as guides. For example, frameworks like the NIST Cybersecurity Framework (CSF) and the NIST AI Risk Management Framework (AI RMF), along with Zero Trust principles, offer a structured, risk-based way to think about scaling AI responsibly. Zero Trust helps define how systems and data should be protected, while NIST frameworks provide a broader governance structure to help ensure AI is secure, reliable, ethical, and aligned with emerging regulations.

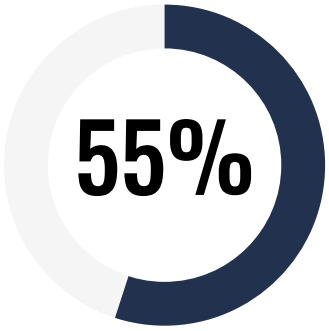
WHERE AVERTIUM CAN HELP

Frameworks like the NIST AI Risk Management Framework are only effective if organizations can apply them in the context of their own environment. That's where many teams struggle to understand what the framework means for their data, their risk profile, and their business priorities. Avertium's NIST AI RMF assessment helps organizations translate frameworks into practical, actionable steps. Whether you're developing AI in-house, integrating third-party models, or scaling AI across business units, the assessment provides a clear, business-aligned path to responsible and secure AI adoption. This equips organizations with:

- Clear insight into current AI-related risks, including shadow or unmanaged AI use.
- A better balance between innovation, productivity, and risk management.
- Alignment with stakeholder expectations and existing business operations.
- A concrete understanding of foundational gaps and remediation priorities.
- The ability to align AI governance with broader regulatory and security frameworks such as NIST CSF, HIPAA, HITRUST, or PCI.

2 TECHNICAL ENABLEMENT: CONFIGURING YOUR ENVIRONMENT FOR AI

Technical enablement makes AI ambition operational and trustworthy, ensuring systems are secure, compliant, accurate, and reliable as they scale. For organizations leveraging platforms like Microsoft, tools such as Copilot are already woven into everyday workflows, helping teams collaborate more efficiently and work faster across familiar environments. But unlocking that potential responsibly requires more than simply turning features on. It demands strategic alignment, technical readiness, and organizational preparedness to ensure AI operates within clear boundaries and earns user trust.



55% of enterprise IT security leaders say they're not fully confident they have the necessary guardrails for deploying AI agents.[4]

For organizations operating in Microsoft environments, readiness often comes down to understanding whether the technical foundations already in place are strong enough to support AI safely at scale. Many teams sense there are gaps, but they lack a clear picture of where those gaps are, how significant they might be, or what to address first.

WHERE AVERTIUM CAN HELP

Avertium's AI readiness assessments, including offerings like the Microsoft Copilot Readiness Assessment, evaluate the strength and maturity of security and identity controls across the environment, helping organizations understand their current technology state before rolling out AI solutions more broadly. During these assessments, Avertium:

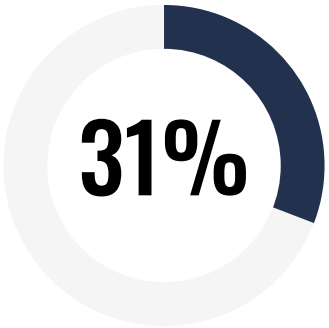
- Reviews security and maturity signals.
- Maps findings against multiple industry frameworks to highlight risk and readiness gaps.
- Identifies practical remediation steps aligned to organizational priorities.
- Builds a clear, phased roadmap for implementation and improvement.

This equips organizations with the confidence they need to scale AI responsibly and securely.

3

DATA-CENTRIC CONTROLS: ESTABLISHING TRUST AT THE SOURCE

As data volumes and generative AI applications grow at an unprecedented pace, the potential for risk grows alongside them. That makes strong data protection and governance essential. By implementing data-centric controls – such as role-based access and least-privilege permissions – and prioritizing the identification and classification of sensitive information, organizations can adopt AI securely and in a way that supports compliance rather than undermines it.



Only 31% of organizations have a clear, fully implemented data governance strategy.[5]

Implementing data-centric controls shifts the focus from simply securing systems to securing the data itself wherever it lives and however it's accessed. Instead of assuming trust based on location or application, having these controls ensures that access, usage, and sharing are governed by the sensitivity of the data and the role of the person or system interacting with it.

HOW AVERTIUM CAN HELP

Tools like Microsoft Purview provide a unified approach to data governance, compliance, and risk management, helping organizations understand their data, apply consistent protections, and ensure responsible data handling as AI becomes more embedded across the business. Avertium helps by putting technology into practice for organizations, building data-centric controls gradually. Avertium starts with foundational governance and expands into more advanced protections as maturity grows. This helps organizations:

- Identify and classify sensitive data.
- Apply labeling and protection policies aligned to business and regulatory needs.
- Reduce the risk of data loss or oversharing.
- Detect and respond to insider and internal data risks.
- Strengthen overall data security posture to support safe AI adoption.

The goal is to make adoption manageable by building a trusted data foundation that enables AI to operate responsibly without compromising productivity or compliance – all while allowing organizations to mature incrementally over time.

BRINGING IT ALL TOGETHER

When governance is clear, technical foundations are sound, and data is protected at the source, organizations can adopt AI with confidence instead of hesitation. With this foundation established, organizations are ready for the next phase of defining purpose-driven AI use cases and shaping the guardrails that ensure agents operate with intention and accountability.

PURPOSE-DRIVEN AI: DEFINING PRIORITIZED USE CASES AND NAVIGATING ADOPTION

As organizations begin laying the groundwork for AI, the natural next question becomes: Where do we start?

While AI's potential seems endless, the smartest path forward is to start small. More specifically, to start purposefully. The organizations seeing real impact from AI are the ones beginning with a single, narrow, well-defined use case that ties directly to a business priority. The reality is many AI projects fail not because the technology wasn't capable, but because teams began piloting before they could articulate what problem they were trying to solve. If you want your AI program to succeed, start small and narrow.

STEP 1

CHOOSE A USE CASE WITH A CLEAR BUSINESS OUTCOME

The strongest AI initiatives begin with a real, felt problem. This might be a care team overwhelmed by manual patient-intake questionnaires, clinicians spending too much time summarizing visit notes, or staff stuck re-entering the same information across disconnected systems. Whatever the catalyst, start by articulating the outcome instead of focusing on the technology.

Ask yourself:

- What outcome do we want to improve?
- Who benefits from solving this?
- What are the ramifications of not addressing the problem?
- How will we measure success?

If those answers aren't clear, the use case isn't ready. A good rule of thumb is if you can't explain the value of the use case without using the words "AI," it's not purpose-driven yet.

PRACTICAL EXAMPLES OF IMPACTFUL EARLY USE CASES

As a cybersecurity MSSP, Avertium has developed early high-value use cases in its own security operations center (SOC), such as:

» NATURAL LANGUAGE QUERY GENERATION

Describe what you want in plain English; let AI write the detection logic.

» ALERT ENRICHMENT

Automatically pull device history, user behavior, or geolocation context into alerts.

» LOW-LEVEL ALERT TRIAGE

Have AI sort, correlate, and suppress routine events so analysts can focus on real risk.

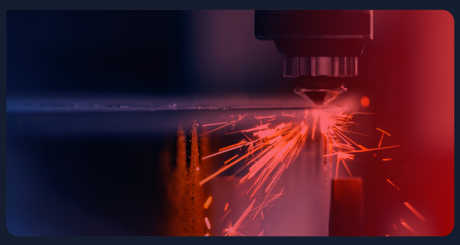
» INCIDENT SUMMARIZATION

Let AI condense long investigative threads into quick-read narratives.

» DRAFTING REPETITIVE PLAYBOOKS

Use AI for the first pass, then refine with human expertise.

But use cases will look different for everyone. Organizations in healthcare, manufacturing, and financial services are prioritizing different kinds of outcome-driven scenarios. The following are some examples.



HEALTHCARE:

Reducing administrative burden and accelerating clinical decisions

Clinical note summarization
Auto condense long physician notes or intake narratives to speed chart reviews.

Patient message triage
Sort and categorize patient portal messages based on urgency, symptoms, or routing rules.

Claims denial pattern detection
Identify recurring denial reasons and recommend optimized appeal language.

MANUFACTURING:

Improving operational efficiency and quality throughput

Quality control defect classification
Turn free text defect logs or operator notes into structured categories for trend analysis.

Maintenance report summarization
Summarize multi-step maintenance logs into concise updates for shift handoffs.

Inventory discrepancy explanation
Analyze sensor data, ERP logs, and operator notes to automatically explain mismatches.

FINANCIAL SERVICES:

Streamlining casework and compliance-heavy workflows

Loan document classification
Auto sort uploaded paystubs, tax documents, and bank statements into required categories.

Customer inquiry summarization
Convert lengthy client communications into quick-read summaries for advisors or service reps.

Transaction narrative enrichment
Translate cryptic transaction strings into human readable descriptions to speed reviews.

STEP 2

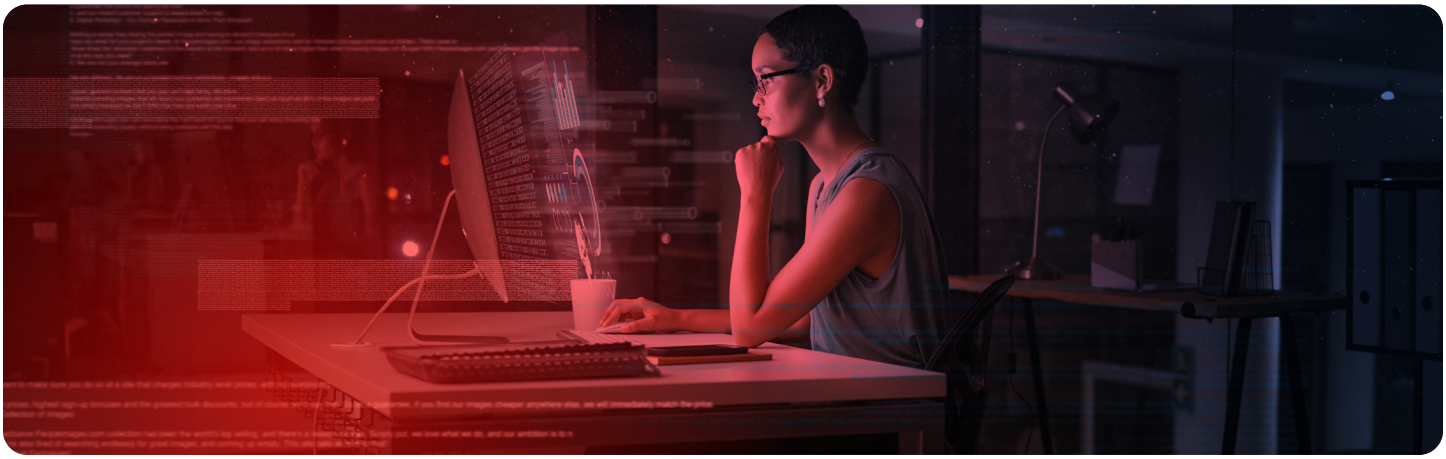
VALIDATE THE USE CASE THROUGH A GOVERNANCE LENS

Once a use case is defined, it must be validated against your organization's governance requirements. This ensures the idea is feasible, safe, and aligned with leadership expectations.

Key considerations include:

- Does this use case comply with regulatory and internal policies?
- What is required to create a robust policy to support this use case?
- What data will the AI touch? Is it governed and labeled appropriately?
- What identity and access controls must be enforced?
- Does the use case require a human-in-the-loop, and at what stages?

This step keeps well-intentioned innovation from becoming shadow AI.



STEP 3

ASSESS THE RISKS AND ESTABLISH GUARDRAILS

Every AI use case introduces new interactions, decisions, and potential consequences. Before implementing, determine the appropriate guardrails based on impact and sensitivity.

This includes:

- Defining what the AI *can* and *cannot* do.
- Establishing oversight and escalation paths.
- Determining where humans validate or overrule AI outputs.
- Identifying any conditions that would pause or restrict automation.

What may initially seem like a limitation is, in fact, a catalyst for creativity – fostering innovation by ensuring adoption is safe, responsible, and sustainable.

STEP 4

PREPARE THE NECESSARY DATA FOR AI

AI is only as effective as the data behind it. Before launching a use case, validate that the appropriate datasets are:

- Accurate
- Labeled
- Governed
- Accessible within proper security boundaries

Poor data hygiene is one of the fastest ways an AI use case can fail. Preparing the data up front ensures the AI performs reliably and reduces the likelihood of incorrect or misleading outputs.

STEP 5

START SMALL, LEARN FAST, AND ITERATE

Begin with the narrowest version of your use case. Start with a constrained workflow, a well-understood dataset, and a small group of analysts. This creates room to learn, adjust, and refine without unnecessary risk.

Once you've proven value and validated safety, scale the use case gradually:

- Widen the scope.
- Add new data inputs.
- Increase automation where appropriate.
- Expand to additional teams.

This step-by-step approach builds confidence and avoids overwhelming your environment – and your staff. The result is AI that feels purposeful as opposed to experimental and delivers measurable ROI instead of unintended complexity.

TAKING YOUR FIRST STEP WITH AI READINESS

Agentic AI will reshape operations, but it will not replace the fundamentals of sound security architecture. The need for human judgment, contextual reasoning, or ethical decision-making will persist. And one truth remains unchanged amid all this transformation: Organizations still need clear visibility into their attack surface. The risks of yesterday are still the risks of today, but more complex. Misconfigurations, over-permissioned identities, unmanaged assets, and blind spots don't disappear simply because AI enters the picture. If anything, they matter more.

That's why AI readiness is the key differentiator between organizations that thrive in the agentic era and those that struggle to translate potential into practice. It means governing your data, closing security gaps, and preparing your people to work responsibly alongside AI. It means putting guardrails in place to keep agents aligned to your mission and prioritizing purpose-driven use cases that deliver real value – not unnecessary complexity. When these elements come together, AI becomes a catalyst for true transformation.

For organizations looking for a partner to guide them through these critical readiness steps, Avertium helps build a strong, AI-ready foundation with clarity, structure, and confidence. From fortifying governance frameworks and securing identities and data to ensuring safe platform configurations and empowering teams to define and implement high-impact use cases, Avertium is your trusted partner at every stage of the journey.

LEARN MORE

[1] ServiceNow Sets New Standard for Fully Autonomous IT, Envisioning a Zero Downtime, Zero Outage Future With Agentic AI, 2025
[2] Beware of Double Agents: How AI can Fortify – or Fracture – Your Cybersecurity, IDC
[3] New Research Finds Only 25 Percent of Organizations Report a Fully Implemented AI Governance Program
[4] Salesforce Research: IT Security Stats for 2025, Salesforce
[5] Dresner Advisory Publishes 2024 Data and Analytics Governance Market Study, PR Newswire

AVERTIUM®

ASSES. DESIGN. PROTECT.