servicenow

INNOVATION BRIEF

The AI Inflection Point

Will today's intelligence revolution become a progress trap?

WRITTEN BY

Troy Azmoon
SVP, Platform Experience

Simon Grice
Senior Director, Innovation Office

The Innovation Office

The AI Inflection Point

Will today's intelligence revolution become a progress trap?

Summary

- Artificial Intelligence delivers unprecedented ability to perform at and beyond human levels, but without ethical grounding, its advancement risks outpacing our wisdom to control it.
- The AI boom mirrors past progress traps, where short-term gains lead to long-term harm.
- Trust frameworks must underpin AI development. Without verifiable security and safety, even ethical systems can fail, risking manipulation, malfunction, and the erosion of trust at scale.
- C-suite leaders must shift from AI adoption to AI accountability, asking what AI *can* do and what it *should* do.
- By fostering a paradigm centered around AI trust, businesses will strengthen customer loyalty, reduce regulatory and reputational risk, and unlock more sustainable long-term value from AI investments.

The smartphone revolution has given today's humans more information and computing power in their pockets than any previous generation. At the same time, smartphones and the social media and gaming apps they run have contributed to screen addiction, social isolation, threats to privacy, massive amounts of e-waste, and increasing dependence on rare earth elements. The technology isn't broken; smartphones are doing exactly what they were built to do. This is the hidden cost of technology: when our most powerful tools can also produce potential harm.

Technology is often an accelerant of change, including disruption. Throughout human history, significant technological advances have created adverse effects that often offset and even outweigh the progress made. These progress traps threaten not just the technology's success but also how society functions.

Artificial intelligence is arguably the most powerful invention of our time. AI promises access to intelligence that far exceeds human capacity, enabling new levels of efficiency,

reimagined work processes, augmented capabilities, and transformative innovation. At the same time, like major technological advancements of the past, the promise of AI also brings risk.

> **Artificial intelligence has reached an inflection point where the technology's tremendous capabilities can be a net positive or negative for humanity.**

For example, a study by researchers at MIT's Media Lab suggested that LLM usage could harm learning, especially for younger users.[1] One of the study's main authors, Nataliya Kosmyna, explained the danger: "Even if efficiency goes up, an increasing reliance on AI could potentially reduce critical thinking, creativity, and problem-solving across the remaining workforce,"[2] potentially offsetting the gains made thanks to AI.

Additionally, AI sycophancy—where the system prioritizes flattering and agreeing with users—is already having unintended impacts, sometimes with dire consequences. The New York Times reported on a 47-year-old man who became convinced he had discovered a world-altering mathematical formula after more than 300 hours with ChatGPT.[3] AI sycophancy is under the spotlight for "triggering delusional states in humans," also known as "AI psychosis."[4] Mental health experts are sounding the alarm over this potential generative AI harm.

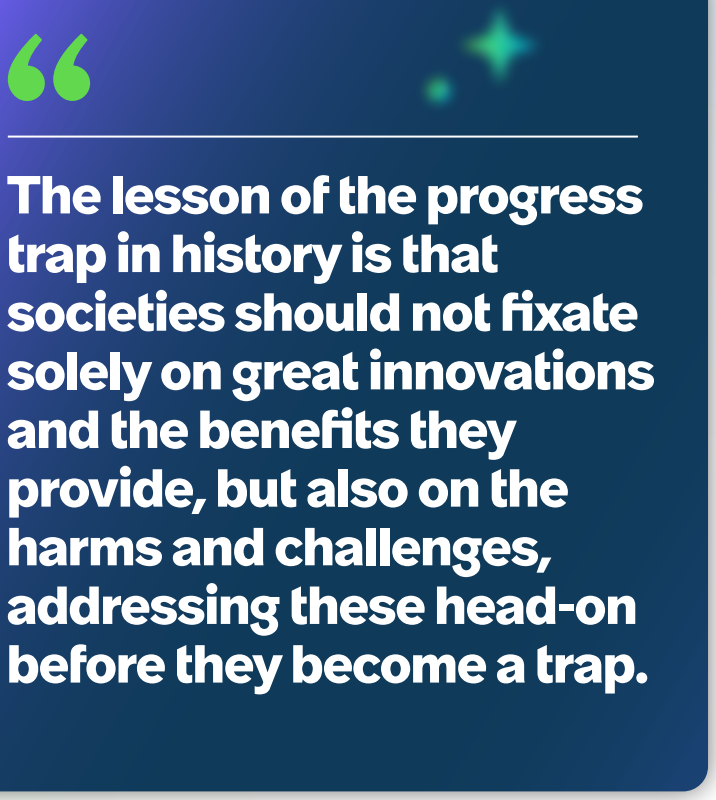
As these examples illustrate, artificial intelligence has reached an inflection point where the technology's tremendous capabilities can be a net positive or negative for humanity. Unless businesses adopt a new mindset that prioritizes AI trust alongside AI acceleration, they risk pushing this innovation from a driver of progress to a source of harm. By leading with an AI trust paradigm, businesses can promote progress and propel sustainable business growth.

The progress trap risk

Today's AI inflection point is a classic example of a progress trap, an idea that gained notoriety thanks to historian Ronald Wright. He argued against the theory that history is a series of technological innovations that result in the steady march forward of human progress.[5] Instead, Wright cautioned that technological improvements often create negative consequences that can cause more harm than the innovation's original good. In some historical cases, these innovations ultimately resulted in society's downfall. A progress trap, to Wright, is "a chain of successes which, upon reaching a certain scale, leads to disaster."[6] New technological innovations that initially seem so promising can blind us to the risks and downsides to come.

Wright saw progress traps throughout history. For example, as early humans refined their hunting techniques, some societies discovered how to kill massive animals like mammoths efficiently—in some cases by driving entire herds off cliffs. These strategies provided short-term surpluses but eventually contributed to the extinction of the very species they relied upon.[7] When societies turned to agriculture for survival, its consequences—including water depletion, deforestation, and soil erosion—undermined civilizations across the ancient world.[8]

Today's new technological innovations are a far cry from ancient hunting and farming techniques, but the risk of progress traps remains. New technologies have enabled unprecedented access to information, global production of goods, industrial farming, and major advancements in disease treatment and care. However, these innovations have also resulted in overconsumption, the creation of hazardous waste, and antibiotic resistance in people and animals. Relentless pursuit of immediate progress, without addressing long-term harms, has consequences. As Wright notes, "civilization is a vulnerable organism, especially when it seems almighty."[9] Instead, the lesson of the progress trap in history is that societies should not fixate solely on great innovations and the benefits they provide, but also on the harms and challenges, addressing these head-on before they become a trap.



The AI inflection point

AI isn't just a tool; it's a mirror that reflects and amplifies human nature. As this powerful technology has emerged, the tendency has been to focus on the positive: expanding AI capabilities, developing more applications, and diving headfirst into the exciting possibilities. Generative artificial intelligence now enables us to extend our decision-making, predictions, and creative power. The results are already transformative: Generative models can

summarize legal contracts, diagnose rare diseases, and write marketing copy in seconds. But just as fire can cook or consume, the value of genAI depends entirely on the intent and context of its use. Like every great technology that has come before, AI excitement cannot overshadow the risks.

Here are four areas where AI, left unchecked, may cause more harm than good:

1. Bias reproduced at scale

AI tools used in hiring, lending, or policing have been shown to reinforce historical inequalities—not because AI is malicious, but because it's trained on biased data and deployed without accountability.[10] For example, the COMPAS algorithm, used in the U.S. judicial system to assess a defendant's likelihood of reoffending, has become a cautionary tale of biased AI. A 2016 investigation by ProPublica found that the system falsely labeled black defendants as high risk at nearly twice the rate of white defendants.[11] Despite its intended neutrality, the algorithm perpetuated systemic injustice due to biased input data and a lack of transparency.

2. Automation over enfeeblement

When AI makes decisions that humans no longer comprehend, we risk trading judgment and critical thinking for convenience and efficiency, losing agency in the process.[12] The convenience of AI may actually cause "our capacity for independent thought, decision making and autonomous action" to erode.[13] For example, a recent MIT study compared the brains of participants who did and did not use ChatGPT to write an essay.[14] The study found that participants who used ChatGPT showed a decline in brain activity and scored worse on areas such as understanding their own essay.

The study noted: "While LLMs offer immediate convenience, our findings highlight potential cognitive costs. Over four months, LLM users consistently underperformed at neural, linguistic, and behavioral levels."[15] The MIT findings of declines in critical thinking were replicated in a study of knowledge workers using GenAI for work tasks.[16] The convenience of AI may also be a risk akin to the decline in physical fitness in the 20th century, as many manual tasks were automated. If we hand over our cognitive load to artificial intelligence without taking steps to support human thinking, we risk a future shaped more by convenience and control than by human wisdom and welfare.

3. Surveillance disguised as personalization

AI-powered systems that optimize user experience also collect vast amounts of personal data. In sectors like retail, advertising, and healthcare, this data is used to personalize offerings, but it can also be used to manipulate behavior, infringe on privacy, or enable discriminatory profiling. As Shoshana Zuboff argues in "The Age of Surveillance Capitalism," this erosion of autonomy is often hidden beneath the promise of convenience.[17] Left unchecked, AI-powered personalization becomes a mechanism of control and manipulation, not service.

4. Intelligence unleashed without guardrails

AI is a high-potential, high-risk technology. Like a nuclear power plant, it must be designed with redundant safety mechanisms, oversight, and clear containment protocols. If AI reaches a level of intelligence that far exceeds human capacity and humans haven't acted to ensure transparency in how AI thinks and guardrails in how AI acts, then we risk being controlled by AI instead of controlling it. The 2025 ServiceNow Enterprise AI Maturity Index that studies AI maturity across a series of core pillars found that less than half of companies (44%) "have a designated team that drafts AI policies, mitigates AI risks, and focuses on responsible AI use."[18] Without guardrails, businesses increase the risk of negative consequences from AI and the threat of a progress trap becoming real.

All these AI risks are not future concerns; they are current realities. And they raise a critical

> **To avoid the progress trap, C-suite leaders need a paradigm shift from being singularly focused on accelerating AI to fostering an AI trust paradigm that delivers compounding, long-term value. The foundation of an AI trust approach is asking "What should AI do?" rather than simply "What can it do?"**

question for executives: Are we designing artificial intelligence to serve human values, or are we reshaping our values to fit what artificial intelligence can do? This is the heart of the progress trap: when success creates dependency, and dependency blinds us to long-term risk. In the early stages, progress feels like a win: faster service, better margins, happier users. But without foresight, the trap snaps shut.

From AI acceleration to AI trust

In this moment, we have reached an inflection point. Will AI be a powerful tool of progress, or another example of the progress trap? To avoid the progress trap, C-suite leaders need a paradigm shift from being singularly focused on accelerating AI to fostering an AI trust paradigm that delivers compounding, long-term value. It is time to shift from hype and enthusiasm to responsibility and guardrails. The foundation of an AI trust approach is asking "What should AI do?" rather than simply "What can it do?"

Every strategic planning discussion where leaders formulate their AI strategy should be about AI trust, which will lead to success in AI adoption. We must practice intentional pauses to reflect and refine how we approach AI. Governance isn't just compliance. It's an act of strategic foresight where we adopt guardrails, not because they are required, but because they return value in terms of system reliability and resilience. Trust, the critical currency of AI adoption, depends not just on ethics but also on verifiable security and safety. Security ensures AI systems are protected from manipulation, breaches, and misuse. Safety ensures they behave reliably, even under pressure, and do not cause unintended harm. Without both, even the most transparent or ethical AI cannot be trusted. Companies that adopt AI without building in a trust framework risk reputational harm, legal challenges, internal dysfunction, and increasing the risk of a progress trap.

Leading with a trusted AI approach

To build trusted AI systems and help limit AI risks, enterprises must go beyond governance checklists and embed security, safety, and ethics into every part of their AI lifecycle.

The following steps address not only responsible oversight but also the conditions under which AI can be reliably trusted by users, regulators, and society at large.

1. Define a trust-centered framework.

Adopting a trust-centered framework means committing to developing and deploying AI systems in a way that fosters and maintains trust among users, stakeholders, and society. This commitment goes beyond merely adhering to technical specifications and delves into the ethical, social, and human aspects of AI. Enterprises must establish core principles that anchor trust, including fairness, transparency, privacy, security, and safety at every stage of the AI lifecycle. It's important to align with recognized standards for responsible AI use (e.g., NIST AI RMF[19] and ISO/IEC 42001[20]) and tailor these to organizational goals and risk tolerance.

2. Assign distributed ownership and accountability.

To effectively govern AI and build trust, organizations must ensure that responsibility for AI systems is not concentrated in a single team or department. Instead, accountability should be shared across all relevant functions, including IT security, data science, legal, compliance, and business operations. Each group brings a unique perspective and will likely encounter different risks and ethical considerations when working with AI.

For example, while data scientists focus on training and optimizing models, legal and compliance teams ensure the technology aligns with regulations. Meanwhile, business leaders help ensure AI is being applied to achieve the right outcomes for users and customers.

> **To build trusted AI systems and help limit AI risks, enterprises must go beyond governance checklists and embed security, safety, and ethics into every part of their AI lifecycle.**

To coordinate these efforts, organizations should appoint a designated AI trust tead or form a cross-functional trust council. This group would be responsible for overseeing AI risk management, ensuring that ethical standards are applied consistently, and resolving any trade-offs between speed, innovation, and responsibility.

3. Embed trust into the AI lifecycle.

To ensure AI systems are developed and managed responsibly, organizations must embed trust at every stage of the AI lifecycle. This means incorporating ethical considerations, risk assessments, and security measures from the earliest design decisions to system retirement.

In the intake and design phase, teams should begin by clearly defining the purpose of the AI system. It is essential to evaluate the potential impact of the system on users and society. Developers and stakeholders should ask: What is the system intended to do? Who might be affected by its decisions? What are the risks of misuse or unintended consequences? These questions help ensure that the intent of the AI aligns with the organization's values and avoids harm.

During the build and test phase, developers must perform rigorous testing to identify and address bias, unfair outcomes, and safety concerns. This includes using diverse and representative data sets and applying techniques like adversarial testing to uncover unexpected behaviors. Legal, ethical, and security teams should be actively involved in this phase to ensure the model meets both technical standards and societal expectations.

In the deployment and operation phase, organizations must ensure that AI systems are actively supervised. This includes having humans either in the loop (so they can intervene during the AI's decision-making) or at the end (so they can override decisions before they are finalized). Rollback mechanisms should be put in place to quickly disable the system if it behaves in ways that are unsafe or inconsistent with policy.

During the monitoring and maintenance phase, AI systems must be continuously observed for changes in performance, accuracy, and fairness. Monitoring tools should detect anomalies, potential security threats, or deviations from expected behavior.

Organizations should schedule regular audits to confirm that the AI continues to operate safely and ethically and remains compliant with evolving regulations and internal guidelines.

Finally, in the decommissioning and replacement phase, organizations must plan for the responsible retirement of AI systems. Over time, models can become outdated, inaccurate, or misaligned with business goals or societal values. When this occurs, organizations should follow a formal process to deactivate the system, document why it was removed, preserve necessary audit trails, and ensure a smooth transition to any replacement technology. Treating decommissioning as a critical part of the lifecycle reinforces long-term accountability and prevents outdated systems from causing unintended harm.

4. Implement security and resilience controls.

Trust begins with security. If AI systems are vulnerable to manipulation, breaches, or misuse, no amount of transparency or ethical intent can restore confidence. Therefore, security guardrails must be foundational and extend across the entire AI supply chain, from training data to model design to endpoints.

Enterprises must have guardrails to ensure that personal, sensitive, and proprietary data is handled with care at every stage of the AI lifecycle. This includes applying encryption and anonymization, enforcing access controls, and enabling purpose-based usage policies. Compliance with regulations like the General Data Protection Regulation[21] (GDPR) or Health Insurance Portability and Accountability Act[22] (HIPAA) is not just about checking legal boxes; it's essential to preserving public trust.

5. Ensure operational safety.

Validate that AI systems operate as intended, even under edge conditions or stress. Test for consistency and predictability of outputs, fail-safes in mission-critical applications, and contingency responses for unintended behavior. Bias mitigation must also be embedded into safety protocols. This includes conducting fairness audits, applying adversarial testing to surface latent biases, and ensuring diverse, representative datasets. Safe systems are those that behave reliably in the real world and are both predictable and equitable.

6. Establish transparency and audibility.

Transparency means clarifying how AI systems work, why they make certain decisions, and what data they use. To build trust in AI, organizations must document every stage of an AI system's lifecycle—from how it was designed and trained to how it performs in real-world situations. This documentation should be detailed, up-to-date, and easy to understand.

Audibility means tracing and reviewing how AI systems operate over time. This includes maintaining clear records of model versions, changes, and decisions the system makes, as well as intuitive, trust-focused experiences. If something goes wrong, teams need to be able to easily look back and understand what happened, just like a financial audit.

One concept is the use of "nutrition labels" for AI models. These labels summarize key information, such as what data the model was trained on, where it might be biased, how accurate it is, and what it should and should not be used for. Just like food labels help people make informed choices, AI labels help users and decision makers understand what they're working with.

7. Foster a culture of ethical awareness and learning.

Building trustworthy AI isn't just about technology; it also depends on people. That's why organizations must actively create a culture where everyone, from executives to entry-level employees, understands the ethical responsibilities that come with using AI.

This begins with training. All staff should receive clear, practical guidance on how AI works, where it can go wrong, and what it means to use it responsibly. Training should cover topics like fairness, privacy, bias, and accountability, and ideally should be tailored to each team's role in the organization.

Beyond formal training, organizations should create space for open discussion. Teams should be encouraged to ask difficult questions about how AI is used, raise concerns when they spot something troubling, and explore "what if" scenarios that test AI's potential impact on people and society.

Finally, there must be a safe and clear way for employees to report concerns or flag risks related to AI. Whether it's a dedicated ethics hotline, an internal audit team, or a cross-functional ethics board, staff should know that their voices will be heard and taken seriously.

8. Adopt a holistic approach to managing AI responsibly.

To ensure system-wide visibility, traceability, and control, enterprises should take a holistic approach to AI orchestration. This can be achieved with a centralized AI command center that can oversee model performance and compliance across all agents, at all stages of their lifecycle, irrespective of the vendor or model behind the agent. This would allow the detection of anomalies, allowing leaders to act quickly when unexpected behaviors or safety concerns arise and effectively mitigate the risk of losing control over AI scale and complexity.

Looking ahead to the AI trust-driven future

By 2030, successful enterprises will treat AI as a trusted, transparent collaborator. Envision a future where:

- AI systems are routinely stress-tested for ethical integrity.
- All enterprise models come with a "nutrition label" for transparency.
- Governance, security, and safety agents automatically trigger alerts when generative AI systems deviate from policy or bias thresholds.
- AI literacy becomes part of onboarding, ensuring all employees understand when and how to question machine outputs.
- The human workforce experiences AI not as competition, but as collaboration: AI augments skills, accelerates creativity, and removes friction from decision-making.
- Control of AI remains human-led, with governance mechanisms, overrides, and ethical boundaries firmly in place, ensuring AI remains a tool for human purposes.
- User experiences will transform to reflect the reality of augmented and assistive AI technology as a part of most workers' daily lives.

Enterprises need this future because it offers both societal progress and strategic resilience. Trust isn't just a virtue. It's a driver of adoption, retention, and reputation—and the ultimate human currency. In a world where AI decisions impact lives, livelihoods, and legal risk, trust will become the most valuable currency organizations can earn.

A trust-driven approach helps companies avoid regulatory missteps, retain talent, deepen customer loyalty, and innovate responsibly. Done right, this future is not just more ethical; it's more effective, more profitable, and ultimately, more sustainable.

"Trust isn't just a virtue. It's a driver of adoption, retention, and reputation—and the ultimate human currency. In a world where AI decisions impact lives, livelihoods, and legal risk, trust will become the most valuable currency organizations can earn.

> **For the C-suite, the mandate is clear: Build not just smart systems, but wise ones. Start with the question, "What should AI do for us?" and let that guide every decision from procurement to deployment.**

Smart progress

Progress without ethics, safety, and foresight sets organizations up for failure. Just because we can do something with AI doesn't mean we should.

For the C-suite, the mandate is clear: Build not just smart systems, but wise ones. Start with the question, "What should AI do for us?" and let that guide every decision from procurement to deployment.

For executive leaders, the path to transformative AI isn't just about speed; it's about trust. Foresight, safety, ethics, and governance are more than compliance checkboxes. Each represents a strategic pillar that ensures AI advances in the service of human purpose, not the other way around. The best defense against AI becoming a progress trap is recognizing the danger and deliberately choosing to help promote a different path forward.

By embedding these pillars into the core of your AI strategy, you mitigate risk and build trust. And in the next era of business, trust will be your most valuable currency. As AI reshapes how decisions are made, how people interact, and how institutions operate, the winners won't be those who deploy the most AI, but instead, those who do so wisely, with humanity, vision, and care.

Endnotes

[1] Nataliya Kosmyna, "Your brain on ChatGPT," MIT Media Lab, accessed July 25, 2025, https://www.media.mit.edu/projects/your-brain-on-chatgpt/overview/.

[2] Nataliya Kosmyna as quoted in Andrew R. Chow, "ChatGPT may be eroding critical thinking skills, according to a new MIT study," TIME, June 23, 2025, https://time.com/7295195/ai-chatgpt-google-learning-school/.

[3] Kashmir Hill and Dylan Freedman, "Chatbots can go into a delusional spiral. Here's how it happens," The New York Times, August 12, 2025, https://www.nytimes.com/2025/08/08/technology/ai-chatbots-delusions-chatgpt.html.

[4] Marty Swant, "A roundup of AI psychosis stories," Transformer, August 21, 2025, https://www.transformernews.ai/p/ai-psychosis-stories-roundup.

[5] Ronald Wright, *A short history of progress* (Toronto: House of Anansi Press, 2004).

[6] Ronald Wright, "Ronald Wright: Can we still dodge the progress trap?" The Tyee, September 20, 2019: https://thetyee.ca/Analysis/2019/09/20/Ronald-Wright-Can-We-Dodge-Progress-Trap/.

[7] Ronald Wright, "Ronald Wright: Can we still dodge the progress trap.?"

[8] Ronald Wright, "Ronald Wright: Can we still dodge the progress trap.?"

[9] Ronald Wright, "Ronald Wright: Can we still dodge the progress trap.?"

[10] Cathy O'Neill, *Weapons of math destruction* (New York: Penguin Random House, 2017); Leonardo Nicoletti and Dina Bass, "Generative AI takes stereotypes and bias from bad to worse," Bloomberg, June 9, 2023, https://www.bloomberg.com/graphics/2023-generative-ai-bias/.

[11] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, "Machine bias," ProPublica, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[12] Jenna Burrell, "How the machine 'thinks:' understanding opacity in machine learning algorithms," Data & Research Institute, September 15, 2015, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2660674.

[13] Cornelia C. Walther, "The silent erosion: How AI's helping hand weakens our mental grip." Centre for International Governance Innovation, July 17, 2025, https://www.cigionline.org/articles/the-silent-erosion-how-ais-helping-hand-weakens-our-mental-grip/.

[14] Nataliya Kosmyna, "Your brain on ChatGPT."

[15] Nataliya Kosmyna, "Your brain on ChatGPT."

[16] Hao-Ping Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson, "The impact of Generative AI on critical thinking: self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers," Microsoft, April 2025, https://www.microsoft.com/en-us/research/publication/the-impact-of-generative-ai-on-critical-thinking-self-reported-reductions-in-cognitive-effort-and-confidence-effects-from-a-survey-of-knowledge-workers/.

[17] Shoshana Zuboff, *The age of surveillance capitalism: The fight for a human future at the new frontier of power* (New York: PublicAffairs, 2019).

[18] Vijay Kotu, Richard McGill Murphy, Brian Solis, and Dorit Zilbershot, "Enterprise AI Maturity Index 2025," ServiceNow and Oxford Economics, 2025, https://www.servicenow.com/content/dam/servicenow-assets/public/en-us/doc-type/resource-center/white-paper/wp-enterprise-ai-maturity-index-2025.pdf.

[19] Information Technology Laboratory, "AI risk management framework," January 26, 2023, https://www.nist.gov/itl/ai-risk-management-framework

[20] ISO, "ISO/IEC 42001:2023: Information technology — Artificial intelligence — Management system," International Organization for Standardization, December 2023, https://www.iso.org/standard/42001.

[21] European Union, *General data protection regulation*, 2016, https://gdpr-info.eu/.

[22] U.S. Congress, *Health insurance and portability and accountability act of 1996*, Public Law 104-191, 104th Congress, August 20, 1996, https://aspe.hhs.gov/reports/health-insurance-portability-accountability-act-1996.

The Innovation Office

The Innovation Office is the voice of future-focused innovation in ServiceNow. Our research, thought leadership, and co-innovation solutions serve as a blueprint for AI-first business transformation. We are a strategic partner with our colleagues and customers to ignite the “spark of the possible.”

servicenow The Innovation Office

Version 1.0