interactions

EBOOK

Secure by Design: Building AI You Can Trust

From regulations to transparency, the principles that keep AI safe and trustworthy

Trust in AI can't come from innovation alone. It comes from embedding security and ethical decisions into every layer of design, every data handling process, and every partnership in the AI supply chain. "Secure by design" is more than a philosophy — it's the foundation for responsible and resilient AI adoption.

Trust in AI can't come from innovation alone. We're seeing that play out on a daily basis. As AI advances at an exhilarating pace, so do concerns around ethics, security, and safety.

General public skepticism around AI can range from data privacy questions to the future of work to more existential debates about creativity, learning, digital reality, and human-AI relationships.

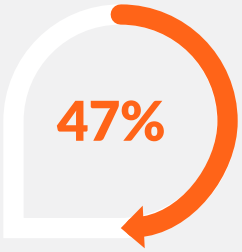
Businesses have long been grappling with data, application, and digital supply chain sprawl due to cloud and big data technologies. But now they must deal with the more unpredictable consequences of AI, including issues that were largely unforeseen just a few years ago: Gen AI- and deep fake phishing; AI hallucinations and biases; model poisoning, evasion, and inversion; prompt injections; and rogue AI that ignores instructions and acts on its own.

There are also practical security issues to manage, including:

- Preventing proprietary and sensitive data from being used to train public models or models that may benefit competitors
- Securing customer and business data on rapidly expanding AI stacks
- Keeping models secure to avoid data leakage, model poisoning, and other issues

To drive AI trust, AI solutions must embed security into every AI modeling and software development step, every data handling process, and every link in the AI lifecycle.

A majority (69%) of business leaders surveyed in Q2 2025 were concerned about AI data privacy, a 60% increase over those who expressed concern just a few months earlier in Q4 2024.[1]



Only 47% of consumers trust AI companies to protect personal data.[2]

The good news is that despite all these shifting paradigms and uncharted territories — fill in your own favorite AI Age cliché — what isn't unprecedented is how to secure data and, therefore, AI," says Bob Steron, SVP CIO and Chief Information Security Officer, at Interactions.

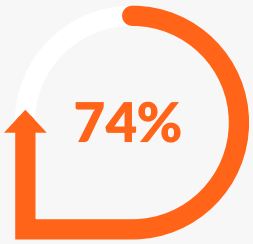
[1]KPMG, "AI Agents Move Beyond Experimentation as Leaders Prepare for Competitive Transformation Within 24 Months," June 24, 2025
[2] Stanford University Human-Centered Artificial Intelligence, The 2025 AI Index Report

AI Security: New(ish) Problems with Proven Solutions

Even though generative AI has taken the spotlight, AI has been a part of our lives for decades. Predictive and classification AI models are used in everything from spam filters to financial auditing, while deterministic models are used for rule-based systems. For example, conversational AI used in contact centers uses predictive models to understand a customer's intent and deterministic models to select responses and perform fixed actions. This behind-the-scenes use of AI has been transformative in business, but relatively invisible to the public.

In contrast, generative AI became a dominant force in just a few years due to its general accessibility via chat interfaces. Now, both consumers and businesses are now racing to keep up with the changes it has inspired. Agentic AI, which can act autonomously to complete a goal by adjusting to new data and changing conditions, will likely prove to be just as influential.

Many of the rising business concerns around AI security are specific to Gen AI: proprietary data being used in models that benefit competitors, leakage of sensitive data into public models, and malicious attacks that seek to reveal PII and internal data. As more AI-driven systems incorporate Gen AI capabilities into the mix, it's understandable that security concerns are increasing.



74% of leaders cited "data privacy and security" as most important to choosing a Gen AI provider, ranking it higher than "technology and expertise."[3]

[3]KPMG, "AI Agents Move Beyond Experimentation as Leaders Prepare for Competitive Transformation Within 24 Months," June 24, 2025

AI, regardless of the type and specific risks, is a technology tool like any other. If used ignorantly or without care, the results may not just be poor but damaging," says Steron. "But at its core, AI is powered by data and accessed by software applications. Securing data and applications are challenges that technology companies like ours have been solving for decades."

The Elements of Secure, Trustworthy AI

For AI adoption to thrive — with purpose, benefit, and curiosity (and not fear, uncertainty, and doubt) — , trustworthy AI solutions must employ not only sound technical controls, such as those that protect customer data during AI training and in operations thereafter, but also be predicated on an AI governance program that embraces both sound principles and values as well as functional frameworks that operationalize those values.

Given the significant impact of AI and its potential for harm, it is critical that organizations identify the values on which their programs will be grounded — and to use these values to find like-minded AI partners that develop ethical, trustworthy solutions.

AI GOVERNANCE: WHERE TO START

As you develop or refine your AI governance programs and frameworks, consider these sources for inspiration and guidance. The Organization for Economic Cooperation and Development (OECD) offers foundational ethical guidelines that include:

- Inclusive growth, sustainable development, and well-being
- Human rights and democratic values, including fairness and privacy
- Transparency and explainability
- Robustness, security, and safety
- Accountability

For the logical next step of operationalizing a values-based AI governance program, adoption of well-known frameworks connotes a level of transparency and due care. Common frameworks include the NIST AI Risk Management Framework (RMF) and ISO 42001: AI Management Systems (AIMS).

Most companies are not building their own AI models and applications. The cost, complexity, and risk are simply too high. Rather, they partner with AI vendors to simplify the complexity and receive market-proven solutions. However, this approach carries some of its own risk, familiar to the business as software supply chain risk. AI may complicate risk with the aforementioned novel threats, but it doesn't foundationally change how vendors should be vetted for security and trustworthiness.

Three key areas can be examined to help determine which companies hold themselves to appropriate standards for building secure, trustworthy AI-powered solutions.

1. Data validation, handling, and protection
2. Secure-by-design AI and software practices
3. Regulatory compliance

Underlying these three areas are two critical tenets: *transparency* and *ethics*. (Sound familiar?) Companies that offer proactive transparency into how they secure their customers' data and their own applications and systems demonstrate stronger integrity and commitment to security than those that do not. This dedication to doing things the right way very likely extends to ethical considerations around AI model bias, fairness, accuracy, and data provenance.

Ultimately," says Steron, "the way technology companies demonstrate (or don't) their ethical position is a strong indicator of how well they might serve as trustworthy, long-term AI partners."

DATA VALIDATION, HANDLING, AND PROTECTION

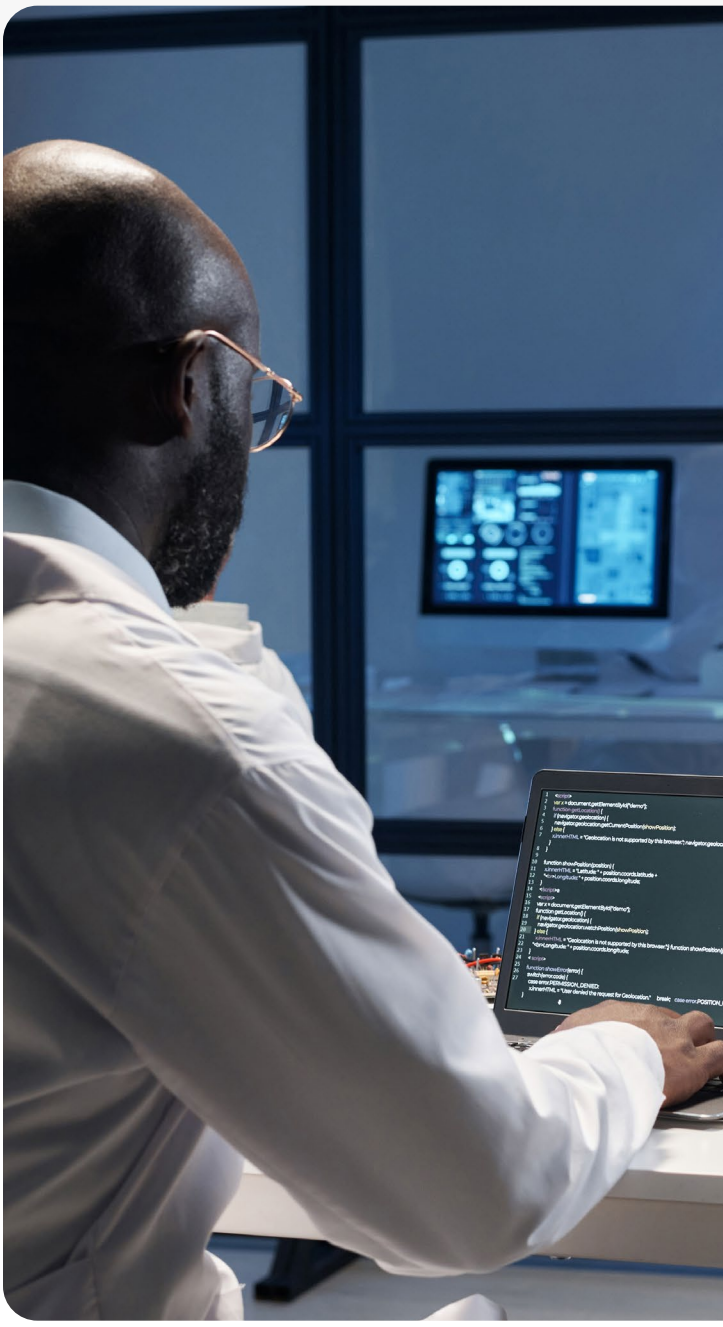
To build truly responsible and trustworthy AI-driven applications, organizations must establish and abide by a meaningful standard for data protection that extends beyond basic compliance. This rigorous approach to data processing during AI model training and AI usage should include:

- Gaining express permission for the use of any data used for AI model training, and not operating AI models in a continuous learning mode by default — especially when sensitive information is being processed.
- Securing or deleting input data after model training.
- Offering the ability to exclude data of certain types from training, such as biometric, PII, and sensitive data.
- Providing an option for single-client, dedicated-use AI models, while educating the client about the tradeoff in results due to a smaller data training set.
- Diligently tracking, monitoring, and documenting data, including the source (to address model poisoning risks), from receipt through destruction. This is a fundamental aspect of GDPR.
- Following, to the letter, client requirements regarding data storage and destruction processes.

Building consumer trust through AI

When it comes to customer care solutions, AI can make interactions more secure, rather than less. For example, a state-of-the-art intelligent virtual assistant (IVA) should be able to:

- Automatically redact PII from call and chat recordings for safer analytics and machine learning.
- Enable fluid transition from live agent to AI when handling PII such as credit card data.
- Automate identity verification consistently and in accordance with the proper regulatory requirements based on the caller's and organization's locations.



SECURE-BY-DESIGN AI AND SOFTWARE PRACTICES

Most companies offering AI solutions are, at their core, software companies that build applications driven by AI. How these applications are conceived, planned, built, and maintained is just as critical to trustworthy AI as safe modeling practices.

Many companies adopt a DevSecOps approach. By embedding security at every stage of the software development lifecycle (SDLC), from coding to post-release, companies should find vulnerabilities earlier — when they're easier to fix — rather than in QA and user acceptance testing (UAT) just before implementation. If SDLC is a challenge for your organization, consider reviewing the OWASP Software Assurance Maturity Model (SAMM), an excellent framework with a very active (and supportive) community of users.

Secure by design (SBD) is a continuation of DevSecOps, elevating security from a feature to be tested to a core business requirement. More than a philosophy, SBD is the foundation for responsible and resilient AI and application development. The moment a project is conceptualized, security must inform fundamental decisions around architecture, authentication, data handling, code bloat, and meeting current and anticipated regulatory demands.

SBD tenets are being pushed both by voluntary efforts, like the U.S. Cyber and Infrastructure Security Agency (CISA) Secure by Design Pledge, and mandated requirements, namely the EU's Cyber Resilience Act. Companies that are working towards the goals outlined by these efforts aren't just demonstrating a commitment to secure software principles, but also readiness for future regulations.

REGULATORY COMPLIANCE

In addition to the operational frameworks mentioned above, a strong sign that a company takes AI security and ethics seriously is embracing, not resenting, rigorous regulatory frameworks like GDPR, CCPA, and both the European Union's EU AI Act and Cyber Resilience Act. In the same vein, Colorado's AI Act, set to become effective on June 30, 2026, will be the strongest AI regulation in the U.S. to date and closely mirrors the EU AI Act by identifying and regulating "high risk" AI systems.

> We love GDPR and are looking forward to clarity provided by the AI Act," says Steron. "We don't see these regulations as legal burdens. They establish the frameworks the industry — and consumers — need to enable new technologies to mature into trustworthy ecosystems."

REGULATIONS ON THE RISE

80.4% of U.S. local policymakers support stricter data privacy rules.[4]

59 AI-related regulations were issued by U.S. federal agencies in 2024, **more than double the number the prior year.**[5]

83% of consumers say personal data protection is one of the most critical factors for businesses to earn their trust.[6]

Adherence to these regulatory constructs, especially when your organization is not mandated to do so, conveys a strong message about your culture with respect to security and ethics. They also create a level playing field for innovation by freeing companies from legal uncertainties and encouraging them to innovate and compete on the quality of their solutions rather than on security or ethical shortcuts. Finally, such regulations help alleviate consumer concerns around privacy.

The ideal AI provider applies the strictest regulations across all products and customers and prepares proactively for phased-in regulations, whether or not those regulations are legally required for all of the vendor's customers. This comprehensive approach doesn't just signal that a company is trustworthy and working above board, Steron notes. "It provides a good indication that they may be a suitable long-term partner that can support your growth, wherever it may take you."

[4] Stanford University Human-Centered Artificial Intelligence, The 2025 AI Index Report

[5] Ibid.

[6] PwC, PwC Voice of the Consumer Survey 2024

Underpinning AI with robust security

Security processes related to AI, data, and development must sit atop a broader, robust cybersecurity program that covers an organization's systems, processes, and people. When performing due diligence on a potential AI partner, seek transparency about:

- Multilayer, redundant cybersecurity processes that prevent single points of failure, including continuous scanning and periodic penetration testing.
- Internal training procedures, including software developer security training, mock phishing exercises, and tracked cybersecurity training on topics like PCI, data security, and privacy.
- Third-party risk management processes (TPRM), i.e., how the vendor vets the third-party AI models and technologies that it uses in its products and operations.
- External audits and certifications, such as PCI-DSS and SOC 2.

Advancing AI and trustworthiness together

AI will keep evolving, faster than feels possible. Agentic AI is poised to further transform how we work, even as society continues to play catch-up with the effects of generative AI. It will also bring its own challenges, like unpredictable behavior or sophisticated cyberattacks.

Ultimately, what will govern the risks of AI are us humans, from the engineers who securely train models, monitor AI output, and build apps, to the leaders who invest in the people, time, and tools needed to build trustworthy AI, to the experts who craft laws to protect consumers and businesses.

> Finding AI partners is as much about integrity and ethics as it is about expertise and innovation. This is a powerful technology, and it's up to all of us to be good stewards by protecting our customers and using AI to deliver better experiences," says Steron.

To learn more about how Interactions is shaping the future of trustworthy AI through security, transparency, ethics, and human oversight, visit our AI Trust Council.

ABOUT INTERACTIONS

Interactions provides Intelligent Virtual Assistants that seamlessly assimilate Conversational AI and human understanding to enable businesses to engage with their customers in highly productive and satisfying conversations. With flexible products and solutions designed to meet the growing demand for unified, omnichannel customer care, Interactions is delivering unprecedented improvements in the customer experience and significant cost savings for some of the largest brands in the world. Founded in 2004, Interactions is headquartered in Franklin, Massachusetts with additional offices worldwide.

For more information about Interactions, contact us:

866.637.9049
interactions.com